

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 04:44:06 ; Search time 20 Seconds
(without alignments)
647.031 Million cell updates/sec

Title: US-10-089-014-1

Perfect score: 1646

Sequence: 1 MSDKSAKIFVAGHRLVGS.....SLRDGLSQTIDYWLKNVCR 312

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	943.5	57.3	321	1 FCL_ECOLI	P32055 escherichia
2	917.5	55.7	314	1 Y4AF_RHISN	P55353 rhizobium s
3	510.0	31.0	312	1 NOLK_AZOCA	P33217 azorhizobiu
4	464.5	28.2	321	1 FCL_MOUSE	P23591 m gdp-fucos
5	463.5	28.2	321	1 FCL_HUMAN	Q13630 homo sapien
6	250.0	15.2	305	1 GALE_METJA	Q57664 methanococc
7	226.0	13.7	334	1 CAPT_STAAR	P39858 staphylococ
8	219.5	13.3	334	1 GALE_BACHD	Q9kdv3 bacillus ha
9	214.0	13.0	361	1 RFBB_SALTY	P26391 salmonella
10	211.0	12.8	329	1 GALE_STRLI	P13226 streptomyce
11	211.0	12.8	361	1 RFBB_SHIFL	P37777 shigella fl
12	209.0	12.7	361	1 RBBI_ECOLI	P37759 escherichia
13	208.0	12.6	338	1 RFPG_HAEIN	P44914 haemophilus
14	206.0	12.5	341	1 ACBB_ACTS5	Q92ae8 actinoplane
15	204.0	12.4	361	1 RBBI_ECOLI	P55293 escherichia
16	201.0	12.2	328	1 STRE_STRGR	P29782 streptomyce
17	193.0	11.7	341	1 RFBB_NEIMA	Q9s642 neisseria m
18	193.0	11.7	350	1 RFBB_RHISN	P55462 rhizobium s
19	192.0	11.7	339	1 GALE_BACSU	P55180 bacillus su
20	190.5	11.6	689	1 GALE_PACTA	P40801 pachysolen
21	190.0	11.5	315	1 SPSS_BACSU	P39630 bacillus su
22	190.0	11.5	699	1 GALE_YEAST	P04397 saccharomyc
23	188.5	11.5	336	1 GALE_YEREN	Q57301 versinia en
24	184.0	11.2	331	1 GALE_LACCA	Q84903 lactobacill
25	183.0	11.1	355	1 RFBB_NEIMB	P55294 neisseria m
26	182.0	11.1	351	1 RFBB_XANCP	P55295 xanthomonas
27	180.0	10.9	350	1 GALE_PEA	Q43070 pisum sativ
28	179.0	10.9	341	1 LPSL_RHIME	O54067 rhizobium m
29	179.0	10.9	355	1 RFPG_ECOLI	P27830 escherichia
30	178.0	10.8	346	1 RFBB_NEIGO	P37761 neisseria g
31	177.0	10.8	329	1 GALE_CORGL	Q45291 corynebacte
32	173.0	10.5	348	1 EXOB_AZOBR	Q59083 azospirillu
33	172.0	10.4	350	1 GAE2_CYATE	O65781 cyanopsais t

RESULT 1
FCL_ECOLI
ID FCL_ECOLI STANDARD; PRT; 321 AA.
AC P32055; P76382;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE GDP-fucose synthetase [includes: GDP-mannose-4-keto-6-D epimerase
(EC 5.1.3.-); GDP-4-keto-6-L-galactose reductase (EC 1.-.-.-)].
GN FCL OR WCAG OR B2052.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RC STRAIN=K12;
RX MEDLINE=95115532; PubMed=7815923;
RA Aoyama K., Haase A.M., Reeves P.R.;
RT "Evidence for effect of random genetic drift on G+C content after
lateral transfer of fucose pathway genes to Escherichia coli K-12.";
RL Mol. Biol. Evol. 11:829-838(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=96326333; PubMed=8759852;
RA Stevenson G., Andrianopoulos K., Hobbs M., Reeves P.R.;
RT "Organization of the Escherichia coli K-12 gene cluster responsible
for production of the extracellular polysaccharide colanic acid.";
RL J. Bacteriol. 178:4885-4893(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Asai H., Baba T., Fujita K., Hayashi K., Inada T.,
Isono K., Kawai H., Kimura S., Kitakawa M., Kitagawa M.,
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
Saito N., Sampei G., Seki Y., Sivasubram S., Tagami J.,
Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98132401; PubMed=9473059;

ALIGNMENTS

RA Andrianopoulos K., Wang L., Reeves P.R.;
 RT "Identification of the fucose synthetase gene in the colanic acid
 RT gene cluster of *Escherichia coli* K-12.";
 RL J. Bacteriol. 180:998-1001(1998).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RC STRAIN=K12;
 RX MEDLINE=99081889; PubMed=9862812;
 RA Somers W.S., Stahl M.L., Sullivan F.X.;
 RT "GDP-fucose synthetase from *Escherichia coli*: structure of a unique
 RT member of the short-chain dehydrogenase/reductase family that
 RT catalyzes two distinct reactions at the same active site.";
 RL Structure 6:1601-1612(1998).
 RN Structure 6:1601-1612(1998).
 CC -|- FUNCTION: TWO STEP NADP-DEPENDENT CONVERSION OF GDP-4-DEHYDRO-6-
 CC DEOXY-D-MANNOSE TO GDP-FUCOSE.
 CC -|- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF THE SLIME POLYSACCHARIDE
 CC COLANIC ACID. SECOND AND THIRD OF THE THREE STEPS IN THE
 CC BIOSYNTHESIS OF GDP-FUCOSE FROM GDP-MANNOSE.
 CC -|- SUBUNIT: HOMODIMER.
 CC -|- SIMILARITY: BELONGS TO THE FUCOSE SYNTHETASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U38473; AAC77843.1;
 CC DR EMBL: AE000295; AAC75113.1;
 CC DR EMBL: D90843; BAA15908.1;
 CC DR PDB: 1BSV; 26-AUG-99.
 CC DR PDB: 1GFS; 12-MAY-00.
 CC DR PDB: 1FXS; 26-AUG-99.
 CC DR EcoGene; EGI1788; fcl.
 CC KW Lipopolysaccharide biosynthesis; Isomerase; Oxidoreductase; NADP;
 CC Multifunctional enzyme; 3D-structure; Complete proteome.
 CC FT CONFLICT 255 256 EL -> DV (IN REF. 1 AND 2).
 CC SEQUENCE 321 AA; 36141 MW; 97077193D79684C7 CRC64;

Query Match 57.3%; Score 943.5; DB 1; Length 321;
 Best Local Similarity 57.6%; Pred. No. 1.4e-72;
 Matches 179; Conservative 56; Mismatches 67; Indels 9; Gaps 1;
 QY 7 KIFVAGHGLVGSIAVRKLOEQGFTNLVLTAEHDLTRQADVESFFSOEKPVYVILAAA 66
 DB 5 RVFIAGHGMVGSIAVRKLOEQGFTNLVLTAEHDLTRQADVESFFSOEKPVYVILAAA 64
 QY 67 KVGGIHANNTPADFIYQNMHESNIHAHQNDVKNLLFLGSSCIYKPAQIPESAL 126
 DB 65 KVGGIHANNTPADFIYQNMHESNIHAHQNDVKNLLFLGSSCIYKPAQIPESAL 124
 QY 127 LTASLEPTNEWAIAKIAGIKTCQAYRIQHGWDAISGMPTNLYGPNDFHNSHVPAL 186
 DB 125 LGQTEPTNEWAIAKIAGIKTCESYRQYGRDYRSMPTNLYGPHDNFPHNSHVPAL 184
 QY 187 MRRFHEAKVNGAEVVMVGTGSPLEFLHVDLADAC-----VFLLDRYSGLEHVN 237
 DB 185 LRRFHEATAQNAQPDVVMVGTGSPLEFLHVDLADAC-----VFLLDRYSGLEHVN 244
 QY 238 IGSQGVVITRELAEVKEVVGEGKLGWCTKPDGTPFKLMDSSKLASLGWTPKYSLRDG 297
 DB 245 VGTGVDTTRELAEVKEVVGEGKLGWCTKPDGTPFKLMDSSKLASLGWTPKYSLRDG 297
 QY 298 LSGTDYWLKN 308
 DB 305 LASTYQWFLN 315

RESULT 2.
 Y4AF_RHISN

ID Y4AF RHISN STANDARD; PRT; 314 AA.
 AC P55353;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 34.7 kDa protein Y4AF.
 GN Y4AF.
 OS *Rhizobium* sp. (strain NGR234).
 OG Plasmid *sym* pNGR234a.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 ON NCBI_TaxID=394;
 RN [1]
 RX MEDLINE=97305956; PubMed=9163424;
 RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
 RT "Molecular basis of symbiosis between *Rhizobium* and legumes.";
 RL Nature 387:394-401(1997).
 CC -|- FUNCTION: PUTATIVE NUCLEOTIDE SUGAR EPIMERASE/DEHYDROGENASE.
 CC -|- COFACTOR: NAD OR NADP.
 CC -|- SIMILARITY: TO NODULATION PROTEIN NOLK, *E. COLI* YEFB AND TO MOUSE
 CC P35B, *M. TUBERCULOSIS* RV1512 AND *Y. ENTEROCOLITICA* WBCJ.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE000064; AAB91603.1;
 CC DR HSP; P32055; 1BSV.
 KW Hypothetical protein; Nodulation; Plasmid; NAD.
 SQ SEQUENCE 314 AA; 34674 MW; 2E571D298AA99110 CRC64;
 Query Match 55.7%; Score 917.5; DB 1; Length 314;
 Best Local Similarity 58.0%; Pred. No. 2.2e-70;
 Matches 177; Conservative 53; Mismatches 74; Indels 1; Gaps 1;
 QY 7 KIFVAGHGLVGSIAVRKLOEQGFTNLVLTAEHDLTRQADVESFFSOEKPVYVILAAA 66
 DB 10 RIWVAGHGMVGSIAVRKLOEQGFTNLVLTAEHDLTRQADVESFFSOEKPVYVILAAA 68
 QY 67 KVGGIHANNTPADFIYQNMHESNIHAHQNDVKNLLFLGSSCIYKPAQIPESAL 126
 DB 69 KVGGIHANNTPADFIYQNMHESNIHAHQNDVKNLLFLGSSCIYKPAQIPESAL 128
 QY 127 LTASLEPTNEWAIAKIAGIKTCQAYRIQHGWDAISGMPTNLYGPNDFHNSHVPAL 186
 DB 129 LTGPLEPTNEWAIAKIAGIKTCQAYRIQHGWDAISGMPTNLYGPNDFHNSHVPAL 188
 QY 187 MRRFHEAKVNGAEVVMVGTGSPLEFLHVDLADACVFLLDRYSGLEHVNIGSQEVTV 246
 DB 189 IRKAHEAKIKDGLCSLWGSPTGTRDFLYSDCDALVFLLKHYSETHINIGSGEISI 248
 QY 247 RELAEVKEVVGEGKLGWCTKPDGTPFKLMDSSKLASLGWTPKYSLRDGLSOTYDYL 306
 DB 249 IELAHIVCRVVGKGDIVFDTSKPDGTPFKLMDSSKLASLGWTPKYSLRDGLSOTYDYL 308
 QY 307 KNVCN 311
 DB 309 SNVAD 313
 RESULT 3
 NOLK_AZOCA
 ID NOLK_AZOCA STANDARD; PRT; 312 AA.
 AC P33217;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Nodulation protein nolK.
GN Azorhizobium caulinodans.
OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hypomicrobium group; Azorhizobium.
OX NCBI_TaxID=7;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORS571;
RX MEDLINE=93113009; PubMed=1472710;
RA Goethals K., Mergaert P., Gao M., Geelen D., van Montagu M.,
RA Holsters M.;
RT "Identification of a new inducible nodulation gene in Azorhizobium
RT caulinodans.";
RL Mol. Plant Microbe Interact. 5:405-411(1992).
CC -|- FUNCTION: PROBABLE NUCLEOTIDE SUGAR EPIMERASE/DEHYDROGENASE.
CC -|- COFACTOR: NAD OR NADP.
CC -|- INDUCTION: BY FLAVANONE NARINGENIN
CC -|- SIMILARITY: BELONGS TO THE FUCOSE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S51942; AAB24744.1; ALT_INIT.
DR HSP; P32055; IBSV.
KW Nodulation; NAD.
FT NP BIND 6 37 NAD/NADP (POTENTIAL).
SQ SEQUENCE 312 AA; 34300 MW; BC2FCB01C87C097C CRC64;

Query Match 31.0%; Score 510; DB 1; Length 312;
Best Local Similarity 35.2%; Pred. No. 6.8e-16;
Matches 107; Conservative 59; Mismatches 136; Indels 2; Gaps 2;

QY 4 KSAKIFVAGHRLVGLVSAIVKLOEQFTNLVLKTHAELDLTROADVSEFFSQKPVVIL 63
Db 3 KGGKLLITGGWVGRNLIAACASRGW-EIIAFTSVDDLRLNAEVEQVIRRLQPDVVH 61
QY 64 AAKVGGIHNNTYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYKFPAPQIPE 123
Db 62 AAGVGGIHNNTYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYKFPACIPELKE 121
QY 124 SALLTASLEPTNEMWAIKAGIKTCQAVRIQHGWDALISGMPNLYGPNDFHPNSHVL 183
Db 122 CDILRGPFEVNTGEGYALAKTVGUKICEYIDKLPNFNFKTLIAICNLVGVGNDFPRSHLL 181
QY 184 PALMRPFHEAKVNGAEVWVGTSPLREPLFHVDDLDACVFLLDYRSLG-EHVNIGSGQ 242
Db 182 PALIEKHKASQGSSESVISWGTARREPFMPAYDFAKIIKALEVELIPSSMNVGVK 241
QY 243 EVTIRELAELVKEVGFEGKLGWDCTKPDGTPKRLMDSSKASLGWTPKVLSDGLSQTY 302
Db 242 DLSVLEYISLVARVIGWGSFVYDLNRPVGMRSKLMIDITHLTALGWVPERSEGGIRSTY 301
QY 303 DWYL 306
Db 302 QYVI 305

RESULT 4
FCL_MOUSE
ID_FCL_MOUSE STANDARD; PRT; 321 AA.
AC P23591;
DT 01-NOV-1991 (Rel. 20, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GDP-fucose synthetase (FX protein) (Red cell NADP(H)-binding protein)
DE (Transplantation antigen P35B) (Tum-P35B antigen) [includes: GDP-
DE mannose-4-keto-6-D epimerase (EC 5.1.3.-); GDP-4-keto-6-L-galactose

reductase (EC 1.-.-.-)].
GN TSTA3 OR TSTAP35B OR P35B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT ASN-139.
RC STRAIN=DBA/2; TISSUE=Maest cells;
RX MEDLINE=90214611; PubMed=2108859;
RA Szikora J.-P., van Pel A., Brichard V., Andre M., van Baren N.,
RA Henry P., de Plaen E., Boon T.;
RT "Structure of the gene of tum- transplantation antigen P35B: presence
RT of a point mutation in the antigenic allele.";
RL EMBO J. 9:1041-1050(1990).
RN [2]
RP REVISIONS.
RA De Plaen E.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: TWO STEP NADP-DEPENDENT CONVERSION OF GDP-4-DEHYDRO-6-
CC DEOXY-D-MANNOSE TO GDP-FUCOSE, INVOLVING AN EPIMERASE AND A
CC REDUCTASE REACTION (BY SIMILARITY).
CC -|- PATHWAY: Conversion of GDP-mannose to GDP-fucose; second step.
CC -|- PATHWAY: Conversion of GDP-mannose to GDP-fucose; third (last)
CC step.
CC -|- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -|- MISCELLANEOUS: MUTAGEN TREATMENT OF P815 TUMOR CELLS PRODUCES TUM-
CC VARIANTS THAT ELICIT A CYTOLYTIC T LYMPHOCYTE RESPONSE (CTL). THE
CC ANTIGENIC ALLELE DIFFERS FROM THE NORMAL ALLELE BY A SINGLE
CC MUTATION IN POSITION 139.
CC -|- SIMILARITY: BELONGS TO THE FUCOSE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53620; CAB94217.1; ALT_SEQ.
DR EMBL; X53621; CAB94217.1; JOINED.
DR EMBL; X53622; CAB94217.1; JOINED.
DR EMBL; X53623; CAB94217.1; JOINED.
DR EMBL; X53624; CAB94217.1; JOINED.
DR EMBL; X53625; CAB94217.1; JOINED.
DR EMBL; X53626; CAB94217.1; JOINED.
DR EMBL; X53627; CAB94217.1; JOINED.
DR EMBL; X53628; CAB94217.1; JOINED.
DR EMBL; M30127; AAA39673.2;
DR EMBL; M30128; AAA39674.1; ALT_FRAME.
DR PIR; S12516; S12516.
DR MGD; MGI:98857; Tstap35b.
KW Isomerase; Oxidoreductase; NADP; Multifunctional enzyme;
KW Tumor antigen; Antigen.
FT NP BIND 9 39 NADP (POTENTIAL).
FT VARIANT 139 139 S -> N (IN TUM- ALLELE).
SQ SEQUENCE 321 AA; 35877 MW; 358D8D6D8F173531 CRC64;

Query Match 28.2%; Score 464.5; DB 1; Length 321;
Best Local Similarity 36.0%; Pred. No. 5e-32;
Matches 113; Conservative 57; Mismatches 129; Indels 15; Gaps 6;

QY 5 SAKIFVAGHRLVGLVSAIVKLOEQF---TNVLKTHAELDLTROADVSEFFSQKPVY 60
Db 7 SMRLVTGSGLVGRAI-QKVADGAGLPGEEWVSVSSKADADLTDAQTALEFKQVQPTH 65
QY 61 VILAAKVGIIHNNTYPADFIGVNLQIQTNVHSAHYEGVKLLFLGSSCIYKFPAPQ 120
Db 66 VIHLAAWVGGLFENIKYNLDLFWKKNVHNDVNLHSAFEGARKVSVCLSTCIPDKTTP 125
QY 121 IPESALLTASLEPTNEMWAIKAGIKTCQAVRIQHGWDALISGMPNLYGPNDFHPNS 180
Db 122 IPESALLTASLEPTNEMWAIKAGIKTCQAVRIQHGWDALISGMPNLYGPNDFHPNS 180

Db 126 IDETMHGPPHNSFGYSYAKRMIDVQNRAYFQHQHCTFTAVIPTNVFGPDNFIENIDG 185
QY 181 HVLPAALMRERHEAKVNGAEVWVGTSPLREFLHVDDLDADACVFLLDYSGLEHV--NI 238
Db 186 HVLPLGLIHKVHLAK-SSDSALVTGCTGKPRQIYSLDIARLFIFWLVREYSEVEPIILSV 244
QY 239 GSGQEVITRELAELVKEVVGFEKGLGWDCTKPGTTPRKMDSSKSLAS----LGWTPKVSU 294
Db 245 GEEDEVSIEKAAEAHVVEAMDENGVEVTFDTSKDGQYKKTASNGKLSYLPDFRFTB---F 301
QY 295 RDGLSQTVDWYLN 308
Db 302 KOAVKETCAWFTDN 315

RESULT 5
FCL_HUMAN
ID FCL_HUMAN STANDARD; PRT; 321 AA.
AC Q13630; Q9UDG7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GDP-fucose synthetase (FX protein) (Red cell NADP(H)-binding protein)
DE [includes: GDP-mannose-4-keto-6-D epimerase (EC 5.1.3.-); GDP-4-keto-6-L-galactose reductase (EC 1.-.-.-)].
GN TSTA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBUNIT.
RC TISSUE=Placenta;
RX MEDLINE=97066899; PubMed=8910301;
RA Tonetti M., Sturla L., Bisso A., Benatti U., De Flora A.;
RT "Synthesis of GDP-L-fucose by the human FX protein.";
RL J. Biol. Chem. 271:27274-27279(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PARTIAL SEQUENCE.
RT TISSUE=Erythrocyte;
RX MEDLINE=95102076; PubMed=7803801;
RA Camardella L., Carratore V., Ciardiello A., Damonte G., Benatti U.,
De Flora A.;
RT "Primary structure of human erythrocyte nicotinamide adenine dinucleotide phosphate (NADP(H))-binding protein FX: identification with the mouse tum-transplantation antigen P35B.";
RL Blood 85:264-267(1995).
CC -!- FUNCTION: TWO STEP NADP-DEPENDENT CONVERSION OF GDP-4-DEHYDRO-6-DEOXY-D-MANNOSE TO GDP-FUCULOSE, INVOLVING AN EPIMERASE AND A REDUCTASE REACTION.
CC -!- PATHWAY: Conversion of GDP-mannose to GDP-fucose; second step.
CC -!- PATHWAY: Conversion of GDP-mannose to GDP-fucose; third (last) step.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE FUCULOSE SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U58766; AAC50786.1; -;
CC DR EMBL; BC001941; AA01941.1; -;
CC DR Genew; HGNC:12390; TSTA3.
CC MIM; 137020; -;

KW Isomerase; Oxidoreductase; NADP; Multifunctional enzyme.
FT NP BIND 5 39 NADP (POTENTIAL)
FT CONFLICT 5 Q -> E (IN REF. 3).
SQ SEQUENCE 321 AA; 35892 MW; 94BB1PF61658007C CRC64;
Query Match 28.2%; Score 463.5; DB 1; Length 321;
Best Local Similarity 35.4%; Pred. No. 6.1e-32;
Matches 111; Conservative 64; Mismatches 124; Indels 15; Gaps 6;
QY 5 SAKIFVAGHGLVGSALVIRKLQEQCF---TNVLKTHAELDLTRQADVESFFSOEKPVY 60
Db 7 SMRIIVTGSGLVGRKI-OKVVADGAGLPGEDWVFSKDDADLTDTATQTRALFEKVQPTH 65
QY 61 VILAAAKVGGIHNANTYPADFIGVNLQIQTNVIHSAYEHGVKGLFLFGSSCIYKFAPOP 120
Db 66 VIHLAAMVGLFRNIKYNLDFWRKNVMDNVLHSAFVGARKVVSCLCIFPDKTYP 125
QY 121 IPESALLTASLEPTNWEYAIKIGIKTCQAYRIHQGWDAISGMTNLNPGDNDHPENS 180
Db 126 IDETMHGPPHNSFGYSYAKRMIDVQNRAYFQHQHCTFTAVIPTNVFGPDNFIENIDG 185
QY 181 HVLPAALMRERHEAKVNGAEVWVGTSPLREFLHVDDLDADACVFLLDYSGLEHV--NI 238
Db 186 HVLPLGLIHKVHLAKSSGS-ALTVMGTGMPRQFIYSLDIAQLFIWLVREYNEVEPIILSV 244
QY 239 GSGQEVITRELAELVKEVVGFEKGLGWDCTKPGTTPRKMDSSKSLAS----LGWTPKVSU 294
Db 245 GEEDEVSIEKAAEAHVVEAMDENGVEVTFDTSKDGQYKKTASNGKLSYLPDFRFTB---F 301
QY 295 RDGLSQTVDWYLN 308
Db 302 KOAVKETCAWFTDN 315

RESULT 6
GALE_METJA
ID GALE_METJA STANDARD; PRT; 305 AA.
AC Q57664;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative UDP-glucose 4-epimerase (EC 5.1.3.2) (Galactowaldenase) (UDP-glucose 4-epimerase).
GN MJ0211.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: UDP-glucose = UDP-galactose.
CC -!- COFACTOR: NAD.
CC -!- PATHWAY: Galactose metabolism; third step.
CC -!- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U58766; AAC50786.1; -;
CC DR EMBL; BC001941; AA01941.1; -;
CC DR Genew; HGNC:12390; TSTA3.
CC MIM; 137020; -;

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC
CC
CC
DR EMBL; U67477; AAB98196.1; -
DR TIGR; MJ0211; -
DR InterPro; IPR001509; Epimerase_Dh.
DR Pfam; PF01370; Epimerase; 1.
KW Hypothetical protein; Isomerase; NAD; Galactose metabolism;
KW Complete proteome.
FT NP BIND 1 32 NAD (POTENTIAL).
SQ SEQUENCE 305 AA; 34542 MW; 2756773CF95D50BD CRC64;

Query Match 15.2%; Score 250; DB 1; Length 305;
Best Local Similarity 24.2%; Pred. No. 6.7e-14;
Matches 78; Conservative 72; Mismatches 130; Indels 42; Gaps 10;

QY 8 IFVAGHRLGVSIVRKLQEQGTNLVLK-----THAELDLTRQADVRS 51
DB 2 ILVTGAGGFIGSHIVDKLIENNYDVIIIDNLTTGNKNNINPKAEFVNADI---RDKDLDE 58
QY 52 FFSQEKPVVILAAAKVGGIHA--NNTYPADFIGVNLQIQTNIHSAIEHGVKKLLFLGS 109
DB 59 KINFKDEVEVIHQAAQINVSNSVENPVYDGD---INVLTGINILEMMRKVDIDKIVFAS 115
QY 110 -SCIYPKFAPQIPESALLTASLEPTNEM--YAIKAGIKTCQAYRIQHGWDALISGMT 166
DB 116 GGAVYGEPNVLPDEN-----HPINPLSPYGLSKYVGEYIKLNLRYLGIEVAILRYS 168
QY 167 NLYGPNDFPHNSHVLPALMRPFHEAKVNGAEVVGWGTGSLRFLHVDLADACVFL 226
DB 169 NVYGERQDPKGE--AGVISIFI-----DKMLKNQSPIIFGNGQTRDFVYGVGVAKANLMA 222
QY 227 LDRYSGLEHVNIGSGEVTIRELAELVKEVWGPGLGWDCTKPDCTPRKLMDSKSLASL 286
DB 223 LNWKN--EIVNIGTGHSTSWELFDIIKHEIGRGEAIYDKPREGEVYRIYLDIKKAESL 280
QY 287 GWTPKVSRLRDLGSLQTYDYLKN 308
DB 281 GWKPEIDLKEGIKRVVNMKNQ 302

RESULT 7
CAP1_STAAU STANDARD; PRT; 334 AA.
AC P39858;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Capi protein.
GN CAP1.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
[1]
SEQUENCE FROM N.A.
RP STRAIN=N;
RX MEDLINE=95050273; PubMed=7961465;
RA Lin W.S., Cunneen T., Lee C.Y.;
RT "sequence analysis and molecular characterization of genes required
RT for the biosynthesis of type 1 capsular polysaccharide in
RT Staphylococcus aureus.";
RL J. Bacteriol. 176:7005-7016(1994).
CC -!- FUNCTION: REQUIRED FOR THE BIOSYNTHESIS OF TYPE 1 CAPSULAR
CC POLYSACCHARIDE.
CC -!- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY. DTDP-GLUCOSE
CC DEHYDRATASE SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC
CC
CC
DR EMBL; U10927; AAA64648.1; -
DR HSSP; P27830; LBXK
DR InterPro; IPR001509; Epimerase_Dh.
DR Pfam; PF01370; Epimerase; 1.
KW Lyase; NAD.
SQ SEQUENCE 334 AA; 37931 MW; 2FDFBF8D4D7796CF CRC64;

Query Match 13.7%; Score 226; DB 1; Length 334;
Best Local Similarity 21.4%; Pred. No. 8.1e-12;
Matches 76; Conservative 71; Mismatches 128; Indels 80; Gaps 11;

QY 7 KIFVAGHRLGVSIVRKLQEQG-----FTNLVLK-----THAELDL 43
DB 2 KILITGTAGTIGSHLAKKLIKQHYVIGVDSINDYVSLSKEDRLKSICKENFTFNKVL 61
QY 44 TRQADVESFSGEKPVVILAAKVGGIHANNTPADFIGVNLQIQTNIHSAIEHGVKK 103
DB 62 ENYDLSKVFVDEQPEVVVNLAQ--AGVRSYIENPRTYIDSNIVGFMNILECSRHFNIQ 120
QY 104 LLFLGSSCIYPKFAPQ-----IPESALLTASLEPTNEMYAIKAGIKTCQAYRIQHG 157
DB 121 LIYASSSSVYGANTSKPFTSDNIDHPLSLYAAIKKSNELMA-----HT 164
QY 158 WDAISGMT-----NLYGPNDFPHNSHVLPALMRPFHEAKVNGAEVVGWGTGSLR 211
DB 165 YSHLYNLPTGLRFFFTVYGPWG--RPDMA-----LFKTKAIVND-QAIDVYVHGNMNR 215
QY 212 EFLHVDLADACVFLDLR-----YSGLEHVNIGSGEVTIRELAELV 253
DB 216 DFTYVDDIIVEAISRLVKKPASPKNKWSGADPDGSSYAPYKVNIGNNSPVLMEFVEAI 275
QY 254 KEVWGPGLGWDCTKPDCTPRKLMDSKL--ASLGTWPKVSLRDLGSLQTYDYLK 307
DB 276 ENKLGKEARKNYMDLQPGDVPETAYNVDDLDFRDFIDPFKPTTIQDGVNKKFVDWYLE 330

RESULT 8
GALE_BACHD STANDARD; PRT; 334 AA.
AC Q9KCV3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-glucose 4-epimerase (EC 5.1.3.2) (Galactowaldenase) (UDP-
DE galactose 4-epimerase).
GN GALE OR BH1108.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
[1]
SEQUENCE FROM N.A.
RP STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- CATALYTIC ACTIVITY: UDP-glucose = UDP-galactose.
CC -!- COFACTOR: NAD.
CC -!- PATHWAY: Galactose metabolism; third step.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

Db 61 DAVAMSRIFAQHPDAVHMLAAE-SHVDRSITGPAAFIETNI-VGTVVLLEARNYSAL 118
 QY 98 EHGKLLFLGSSCIYKPAQPIESALLTASLEPTNEWYATAKIGTKOAY--RIQ 155
 Db 119 NDEKKSKFRPHHISTDEVYGDLPHPDEANNEALPLFTTAYAPSSPYASAKASDHLV 178
 QY 156 HGWDAISGMT-----NLYGPNDFH-PENSHVLPALMRPFHEAKVNGAEVVVWGTS 208
 Db 179 RAMKRTYGLPTIVTNCNNYGP---YHFPEK--LIPVLINALEGKA-----LPIYKGD 228
 QY 209 PLREFLHVDDADACVFLLDYSLGLEHVNIGSQE-----VTIRELA-ELVKEVVGFE 260
 Db 229 QIRDWLYVEDHARALYTVVTEGKAGETYNIGHNEKNKIDVVLITICDLLDEIVPKEKSYR 288
 QY 261 KGLGWDCTKPDGTGPRKMLDSSKLA-SLGWTPKVSRLDGLSQTVDWYLN 308
 Db 289 EQITYVADRPGRHRRYDAIDAKISRELGMKPQETFSGIRKTVWYLAN 337

RESULT 12

RBBI_ECOLI
 ID RBBI_ECOLI STANDARD; PRT; 361 AA.
 AC P3759; P78082;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE dTDP-glucose 4,6-dehydratase (EC 4.2.1.46).
 GN RFB OR RMLB OR B2041.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / WGI;
 RX MEDLINE=94292435; PubMed=7517391;
 RA Stevenson G., Neal B., Liu D., Hobbs M., Packer N.H., Batley M.,
 RA Redmond J.W., Lindquist L., Reeves P.R.;
 RT "Structure of the O antigen of Escherichia coli K-12 and the sequence
 of its rfb gene cluster.";
 RL J. Bacteriol. 176:4144-4156(1994).
 RN [2]
 RP REVISIONS TO 123 AND 250.
 RC STRAIN=K12 / WGI;
 RA Stevenson G.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9037040;
 RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map";
 RL DNA Res. 3:379-392(1996).
 CC -I- CATALYTIC ACTIVITY: dTDP-glucose = dTDP-4-dehydro-6-deoxy-D-
 glucose + H(2)O.
 CC -I- COFACTOR: NAD.

CC -I- PATHWAY: dTDP-L-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
 BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -I- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY. dTDP-GLUCOSE
 DEHYDRATASE SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL; U09876; AAB88398.1; -.
 DR EMBL; AE000294; AAC75102.1; -.
 DR EMBL; D90841; BAA15883.1; -.
 DR EMBL; D90842; BAA15895.1; -.
 DR HSSP; F27830; 1BXX.
 DR EcoGene; EG12412; rfbB.
 DR InterPro; IPR001509; Epimerase_Dh.
 DR Pfam; PF01370; Epimerase_1.
 DR TIGRFAMs; TIGR01181; dTDP_gluc_dehyd; 1.
 KW Lipopolysaccharide biosynthesis; Lyase; NAD; Complete proteome.
 FT NP_BIND 7 13 NAD (POTENTIAL).
 FT CONFLICT 123 123 K -> E (IN REF. 1).
 FT CONFLICT 250 250 G -> V (IN REF. 1).
 SQ SEQUENCE 361 AA; 40558 MW; C3F64643271C14C7 CRC64;
 Query Match 12.7%; Score 209; DB 1; Length 361;
 Best Local Similarity 25.4%; Pred. No. 2.5e-10;
 Matches 89; Conservative 59; Mismatches 139; Indels 64; Gaps 17;
 QY 7 KIFVAGHREGLVGSATVRKL---QEQFTNLVLKTHA-----ELDLTROADV-- 49
 Db 2 KILVTGGAGFICSAVVRHHIINTQSVNVNPKLYAGNRESLADVSDSERVYFERRADICD 61
 QY 50 ----ESFFSQBKPVVYVILAAAKVGGIHNANTYPADFIGVNLQIOTNVHSAVEH----- 99
 Db 62 APAMARIFAQHPDAVHMLAAE-SHVDRSITGPAAFIETNI-VGTVVLLEARNYSALD 119
 QY 100 GVKLLP-----LGSSCIYKPAQPIESALLTASLEPTNEWYATAKIGTKOAY--R 153
 Db 120 SDKKNSFRPHHISTDEVY---GDLPHDPDEVNTEELPLFTTAYAPSSPYASAKASD 176
 QY 154 IQHGWDAISGMT-----NLYGPNDFH-PENSHVLPALMRPFHEAKVNGAEVVVWGCT 206
 Db 177 LVRAWKRTYGLPTIVTNCNNYGP---YHFPEK--LIPVLINALEGKA-----LPIYK 226
 QY 207 GSPLREFLHVDDADACVFLLDYSLGLEHVNIGSQE-----VTIRELA-ELVKEVVG 258
 Db 227 GDQIRDWLYVEDHARALYTVVTEGKAGETYNIGHNEKNKIDVVLITICDLLDEIVPKEKS 286
 QY 259 FEGKLGWDCTKPDGTGPRKMLDSSKLA-SLGWTPKVSRLDGLSQTVDWYLN 308
 Db 287 YREQITYVADRPGRHRRYDAIDAKISRELGMKPQETFSGIRKTVWYLSN 337
 RESULT 13
 RFFG_HAEIN
 ID RFFG_HAEIN STANDARD; PRT; 338 AA.
 AC P44914;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE dTDP-glucose 4,6-dehydratase (EC 4.2.1.46).
 GN RFFG OR H10873.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 05:04:31 ; Search time 531 Seconds
(without alignments)
59.476 Million cell updates/sec

Title: US-10-089-014-1
Perfect score: 1646
Sequence: 1 MSDKSAKIFVAGHRLVGSAL.....SLRDLSQTYDWYLNKVNCR 312

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101233694 residues

Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pap.*
- 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap.*
- 7: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pap.*
- 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pap.*
- 10: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap.*
- 11: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pap.*
- 12: /cgn2_6/prodata/1/pubpaa/US10_PUBCOMB.pap.*
- 13: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pap.*
- 14: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1515	92.0	314	10	US-09-318-271-2 Sequence 2, Appli
2	1241	75.4	310	10	US-09-734-569-148 Sequence 148, App
3	937.5	57.0	321	10	US-09-318-271-4 Sequence 4, Appli
4	671	40.8	162	10	US-09-734-569-16 Sequence 16, Appl
5	463.5	28.2	321	10	US-09-318-271-6 Sequence 6, Appli
6	249	15.1	311	9	US-09-738-626-3862 Sequence 3862, Ap
7	236	14.3	374	9	US-10-168-066-7 Sequence 7, Appli
8	215	13.1	420	9	US-10-028-072-260 Sequence 260, App
9	215	13.1	420	9	US-10-121-049-260 Sequence 260, App
10	215	13.1	420	9	US-10-123-904-260 Sequence 260, App
11	215	13.1	420	9	US-10-140-470-260 Sequence 260, App
12	215	13.1	420	9	US-10-175-746-260 Sequence 260, App
13	215	13.1	420	9	US-10-176-918-260 Sequence 260, App
14	215	13.1	420	9	US-10-176-921-260 Sequence 260, App
15	215	13.1	420	9	US-10-227-884-240 Sequence 240, App
16	215	13.1	420	9	US-10-137-865-260 Sequence 260, App
17	215	13.1	420	9	US-10-140-474-260 Sequence 260, App
18	215	13.1	420	9	US-10-142-431-260 Sequence 260, App
19	215	13.1	420	9	US-10-143-114-260 Sequence 260, App

ALIGNMENTS

RESULT 1

US-09-318-271-2

; Sequence 2, Application US/09318271A

; Patent No. US20020012979A1

; GENERAL INFORMATION:

; APPLICANT: Berry, Alan

; APPLICANT: Running, Jeffrey A.

; APPLICANT: Severson, David K.

; APPLICANT: Burlingame, Richard P.

; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"

; FILE REFERENCE: 3161-24

; CURRENT APPLICATION NUMBER: US/09/318,271A

; CURRENT FILING DATE: 1999-05-25

; EARLIER APPLICATION NUMBER: 60/125,073

; EARLIER FILING DATE: 1999-03-17

; EARLIER APPLICATION NUMBER: 60/125,054

; EARLIER FILING DATE: 1999-03-18

; EARLIER APPLICATION NUMBER: 60/088,549

; EARLIER FILING DATE: 1998-06-08

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 2

; LENGTH: 314

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-318-271-2

Query Match 92.0%; Score 1515; DB 10; Length 314;

Best Local Similarity 92.5%; Pred. No. 5.9e-142;

Matches 298; Conservative 1; Mismatches 5; Indels 18; Gaps 5;

QY 1 MSDKSAKIFVAGHRLVGSALVRKLOEQGFTNLVLKTHAELDLTROADVESFFSQKPVY 60

|||||

1 MSDKSAKIFVAGHRLVGSALVRKLOEQGFTNLVLKTHAELDLTROADVESFFSQKPVY 60

|||||

QY 61 VILAAKVGIIHANNYPADFGVNLQIQTNVIHSAHYGVKKLLFLGSSCIYKPAQP 120

|||||

61 VILAAKVGIIHANNYPADFGVNLQIQTNVIHSAHYGVKKLLFLGSSCIYKPAQP 120

|||||

QY 121 IPESALLTASLEPTNEWYAIKAGIKTCQAVRIQHWDAISGMPTNLGPNDFHPENS 180

|||||

121 IPESALLTASLEPTNEWYAIKAGIKTCQAVRIQHWDAISGMPTNLGPNDFHPENS 180

|||||

;; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
;; FILE REFERENCE: BASP-NAE-1332-99-US
;; CURRENT APPLICATION NUMBER: US/09/734,569
;; PRIOR FILING DATE: 2001-05-24
;; PRIOR FILING DATE: 1999-12-16
;; NUMBER OF SEQ ID NOS: 181
;; SOFTWARE: Patent in Ver. 2.1/WordPerfect 6.1
;; SEQ ID NO 16
;; LENGTH: 162
;; TYPE: PRT
;; ORGANISM: Physcomitrella patens
US-09-734-569-16

Query Match 40.8%; Score 671; DB 10; Length 162;
Best Local Similarity 74.8%; Pred. No. 1.1e-58;
Matches 119; Conservative 25; Mismatches 15; Indels 0; Gaps 0;

QY 150 QAYRIQHGWDAISGMPNTNLYGNDNPHNSHVLPALMRPFHEAKVNGAEVVMVGTGSP 209
Db 1 QAYRLQYNFDAISGMPNTNLYGNDNPHNSHVLPALMRPFHEAKVNGAEVVMVGTGSP 60

QY 210 LREFLHVDLADACVFLDRLYSLGHEVNIHSGOEVITIRELAELVKEVVGFEKLGWDCTK 269
Db 61 PFEFLHVDLADACVFLDRLYSLGHEVNIHSGOEVITIRELAELVKEVVGFEKLGWDCTK 120

QY 270 PDGTPRKLMDSSKSLASLGTWPKVSLRDGLSQTVDWYLNK 308
Db 121 PDGTPRKLMDSSKSLASLGTWPKVSLRDGLSQTVDWYLNK 159

RESULT 5
US-09-318-271-6
;; Sequence 6, Application US/09318271A
;; Patent No. US20020012979A1
;; GENERAL INFORMATION:
;; APPLICANT: Berry, Alan
;; APPLICANT: Running, Jeffrey A.
;; APPLICANT: Severson, David K.
;; APPLICANT: Burlingame, Richard P.
;; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"
;; FILE REFERENCE: 3161-24
;; CURRENT APPLICATION NUMBER: US/09/318,271A
;; CURRENT FILING DATE: 1999-05-25
;; EARLIER APPLICATION NUMBER: 60/125,073
;; EARLIER FILING DATE: 1999-03-17
;; EARLIER APPLICATION NUMBER: 60/125,054
;; EARLIER FILING DATE: 1999-03-18
;; EARLIER APPLICATION NUMBER: 60/088,549
;; EARLIER FILING DATE: 1998-06-08
;; NUMBER OF SEQ ID NOS: 15
;; SOFTWARE: Patent in Ver. 2.0
;; SEQ ID NO 6
;; LENGTH: 321
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-318-271-6

Query Match 28.2%; Score 463.5; DB 10; Length 321;
Best Local Similarity 35.4%; Pred. No. 1e-37;
Matches 111; Conservative 64; Mismatches 124; Indels 15; Gaps 6;

QY 5 SAKIFVAGHGLVGSALVRKLEQGF---TNLVKTHAELDLTROADVSEFFSOEKPVY 60
Db 7 SMRLVTGGSLVGRKAI-QKVADGAGLPGEDVWFVSSKADLTDTAQTRALFEKVPQPH 65

QY 61 VILAAKVGGIHANNTPADFTGNVLQIQTNNVHSAHYHGKLLFLGSSCTYKPFAPQP 120
Db 66 VTHLAAMVGGFLPNIKYNLDLFWKKNVHMNDNLHSAFEVARGKAVSCSLCTCFDPKTTY 125

QY 121 IPESALLTASLEPTNEWYAIAGIKTCQAYRIQHGWDAISGMPNTNLYGNDNPHNS 180

Db 126 IDETHIHGPPHNSNFGSYAKRMIDVQNRAYFOQYGCFTTAVITNVFGPHDNFIEDG 185
QY 181 HVLPALMRPFHEAKVNGAEVVMVGTGSPFLREFLHVDLADACVFLDRLYSLGHEV--NI 238
Db 186 HVLPLIHKVHLAKSSGS-ALTVMGTGNPRQFIYSLDLAQLFIWLVREYNEVEPIILSV 244
QY 239 GSGQEVITIRELAELVKEVVGFEKLGWDCTKPDGTPRKLMDSSKSLAS-----LGTWPKVSL 294
Db 245 GEEDEVSIKEAAEAAYVEAMDPEHGEVTFDTTKSDGQFKKTASNSKLRTYLPDRFETP---F 301
QY 295 RDGLSQTVDWYLNK 308
Db 302 KOAVKETCAWFTDN 315

RESULT 6
US-09-738-626-3862
;; Sequence 3862, Application US/09738626
;; Publication No. US20020197605A1
;; GENERAL INFORMATION:
;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAOKO
;; APPLICANT: SENOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738,626
;; CURRENT FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: Patent in ver. 3.0
;; SEQ ID NO 3862
;; LENGTH: 311
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3862

Query Match 15.1%; Score 249; DB 9; Length 311;
Best Local Similarity 25.3%; Pred. No. 1.8e-16;
Matches 82; Conservative 55; Mismatches 139; Indels 48; Gaps 9;

QY 10 VAGHGLVGSALVRKLEQGFNTLVK-----THAELDLTROADVE 50
Db 5 VTGGAGFTGSHLDLILKEGHEVVVIDNLSSRGRLENLSDAEATGKLTFFVADLL-DVDNF 63

QY 51 SFFSOEKPVYVILAAAKVGGIHANNTPADFTGNVLQIQTNNVHSAHYHGKLLFLGSG- 109
Db 64 EFLGTHKEVTFHLLAAQIDVRH-SVVDPLHDAETHILSTIRIADAAHQGVKRVFTSSG 122

QY 110 SCIYKPKFAPQIPESALITASLEPTNEWYAIAGIKTCQAYRIQHGWDAISGMPNTNLY 169
Db 123 GSIYGESEFPVDE---TVPVDHSP-YAASKVSGEYIYNTFRHLYGLDCSHIAPANVY 177

QY 170 GENDNFHNSHVLPALMRPFHEAKVNGAEVVMVGTGSPFLREFLHVDLADACVFLDRL 229
Db 178 GERQDPHGEAGVAIFALR-----LLGGLDTPKFGDGNTRDYVYVGVVRAFAFLASGE 231

QY 230 YSLGSHVNIHSGOEVITIRELAELVKEVVGFEKLGWDCTKPDGTPRKLMDSSKSLA----- 284
Db 232 IGGGERFNIGTSVETISDRQLHTLVATAAGSK-----DDPEYAPARLGDVPRSAISFK 284

QY 285 ---SLGWTPKVSRLDGLSQTVDY 305

RESULT 9

US-10-121-049-260

; Sequence 260, Application US/10121049

; Publication No. US2003002239A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tamas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330R1C17

; CURRENT APPLICATION NUMBER: US/10/121,049

; CURRENT FILING DATE: 2002-04-12

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 260

; LENGTH: 420

; TYPE: PRT

; ORGANISM: Homo Sapien

; US-10-121-049-260

Query Match

13.1%; Score 215; DB 9; Length 420;

Best Local Similarity 25.3%; Pred. No. 6.6e-13;

Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

QY 1 MSDKSAC-IFVAGHRLVGSIAVRKLOEQGFNTLVKLT-----HAELDLTR 45

Db 83 LSEKDRKRILITGGAGFVGSHTDKLMDGHEVTVVDNFTGKRKNVEHWIGHENFELIN 142

QY 46 QADVESFFSQEPKPVYVILAAAKVGGIHANNTPADFIGVNLQ-IQTNVIHSAYEHGVKK- 103

Db 143 HDVVEPLYEVDQIYHLASPAS-----PPNYMYPNPKTLKNTIGTLMNGLAKR 192

QY 104 ----LLFLGSSCIYPKFAPQIPESALLTASLEP--TNEWYAIAGIKTKQAYRIQHG 157

Db 193 VGARLLLASTSEVYGD--PEVHPQSEDYWGHNPIGPACRYDEGKRVAETMTCYAYMKQSG 250

QY 158 WDAISGMPTNLVGP-----NDNFHPENSHVLPALMRFPHEAKVNGAEVVVWGTGSPLEPREF 213

Db 251 VEVRVARIFTFGRPMHNDG-RVVSNFILQALQ-----GEPLTVYSGSGQTRAF 299

QY 214 LHVDDLADACVFLDLDRYSGLEHVNIGSGOEVTTIRELAELVKEVGFEGKLGW----- 265

Db 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTILEFAQLIKNLVSGSGEIQFLSEAQDDP 358

QY 266 DCTKPDGTTPRKLMDSKSLASLGWTPKVSRLDGLSQTIDWYLNKV 309

Db 359 QKRKPDIKKAKLM-----LGWEPVPLEEGLNKAHYFRKEL 395

RESULT 10

US-10-123-904-260

; Sequence 260, Application US/10123904

; Publication No. US20030022328A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 260
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-123-904-260

Query Match 13.1%; Score 215; DB 9; Length 420;
Best Local Similarity 25.3%; Pred. No. 6.6e-13;
Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

QY 1 MSDKSAC-IFVAGHRLVGSIAVRKLOEQGFNTLVKLT-----HAELDLTR 45
Db 83 LSEKDRKRILITGGAGFVGSHTDKLMDGHEVTVVDNFTGKRKNVEHWIGHENFELIN 142
QY 46 QADVESFFSQEPKPVYVILAAAKVGGIHANNTPADFIGVNLQ-IQTNVIHSAYEHGVKK- 103
Db 143 HDVVEPLYEVDQIYHLASPAS-----PPNYMYPNPKTLKNTIGTLMNGLAKR 192
QY 104 ----LLFLGSSCIYPKFAPQIPESALLTASLEP--TNEWYAIAGIKTKQAYRIQHG 157
Db 193 VGARLLLASTSEVYGD--PEVHPQSEDYWGHNPIGPACRYDEGKRVAETMTCYAYMKQSG 250
QY 158 WDAISGMPTNLVGP-----NDNFHPENSHVLPALMRFPHEAKVNGAEVVVWGTGSPLEPREF 213
Db 251 VEVRVARIFTFGRPMHNDG-RVVSNFILQALQ-----GEPLTVYSGSGQTRAF 299
QY 214 LHVDDLADACVFLDLDRYSGLEHVNIGSGOEVTTIRELAELVKEVGFEGKLGW----- 265
Db 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTILEFAQLIKNLVSGSGEIQFLSEAQDDP 358
QY 266 DCTKPDGTTPRKLMDSKSLASLGWTPKVSRLDGLSQTIDWYLNKV 309
Db 359 QKRKPDIKKAKLM-----LGWEPVPLEEGLNKAHYFRKEL 395

RESULT 11
US-10-140-470-260
; Sequence 260, Application US/10140470
; Publication No. US20030022331A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven

; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 260
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-260

Query Match 13.1%; Score 215; DB 9; Length 420;
Best Local Similarity 25.3%; Pred. No. 6.6e-13;
Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;
QY 1 MSDKSAK-IFVAGHRLVGSIAVRKLEQOGFTNLVLT-----HAELDLTR 45
Db 83 LSEKDRKRILITGGAGFVGSHTDLKMDGHEVTVVDNFFTKRKNVHGHENFELIN 142
QY 46 QADVESFFSQEKPVYVILAAAKVGGIHANNTPADFIGVNLQ-IQTNVIHSAVEHGVKK- 103
Db 143 HDVVEPLYIEVDQIYHLASPAS-----PPNYMYPNFIKLTNTIGTLMNGLAKR 192
QY 104 ----LFLGSSCIYKPAQPIPEPESALLTASLEP--TNEWYAIKTIAGIKTCQAYRIQHG 157
Db 193 VGARLLLASTSEVIGD--PEVHPQSEDYMGVHNPIGPRACYDEGKRVAEETMCTAYMKQEG 250
QY 158 WDAISGMPTNLYGP-----NDNFHPENSHVLPALMRFRHEAKVNGABEVVWGTGSPREF 213
Db 251 VEVRVARIFNTFGFRMHMDNG-RVNSNFILOALQ-----GEPLTVYSGSQTRAF 299
QY 214 LHVDDLDACVFLDLRYSGLEHVNIGSGQEVTTIRELAELVKVEVVGEGKLGW----- 265
Db 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTTLEFAQLIKNLVSGSGEIQFLSEAQDDP 358
QY 266 DCTKPDGTPRKMDSSKSLASLGWTPKVSRLDGLSQTYDWYLVKNV 309
Db 359 QKRKPDIKKAKLM-----LGWEPVVPLEEGLNKAIHYPRKEL 395

RESULT 12
US-10-175-746-260
; Sequence 260, Application US/10175746
; Publication No. US20030027270A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C353

; CURRENT APPLICATION NUMBER: US/10/175,746
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 260
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-746-260

Query Match 13.1%; Score 215; DB 9; Length 420;
Best Local Similarity 25.3%; Pred. No. 6.6e-13;
Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;
QY 1 MSDKSAK-IFVAGHRLVGSIAVRKLEQOGFTNLVLT-----HAELDLTR 45
Db 83 LSEKDRKRILITGGAGFVGSHTDLKMDGHEVTVVDNFFTKRKNVHGHENFELIN 142
QY 46 QADVESFFSQEKPVYVILAAAKVGGIHANNTPADFIGVNLQ-IQTNVIHSAVEHGVKK- 103
Db 143 HDVVEPLYIEVDQIYHLASPAS-----PPNYMYPNFIKLTNTIGTLMNGLAKR 192
QY 104 ----LFLGSSCIYKPAQPIPEPESALLTASLEP--TNEWYAIKTIAGIKTCQAYRIQHG 157
Db 193 VGARLLLASTSEVIGD--PEVHPQSEDYMGVHNPIGPRACYDEGKRVAEETMCTAYMKQEG 250
QY 158 WDAISGMPTNLYGP-----NDNFHPENSHVLPALMRFRHEAKVNGABEVVWGTGSPREF 213
Db 251 VEVRVARIFNTFGFRMHMDNG-RVNSNFILOALQ-----GEPLTVYSGSQTRAF 299
QY 214 LHVDDLDACVFLDLRYSGLEHVNIGSGQEVTTIRELAELVKVEVVGEGKLGW----- 265
Db 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTTLEFAQLIKNLVSGSGEIQFLSEAQDDP 358
QY 266 DCTKPDGTPRKMDSSKSLASLGWTPKVSRLDGLSQTYDWYLVKNV 309
Db 359 QKRKPDIKKAKLM-----LGWEPVVPLEEGLNKAIHYPRKEL 395

RESULT 13
US-10-176-918-260
; Sequence 260, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 260
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-918-260

Query Match 13.1%; Score 215; DB 9; Length 420;
 Best Local Similarity 25.3%; Pred. No. 6.6e-13;
 Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

QY 1 MSDKSAK-IFVAGHRLGVSIAVRKLOEQGFTNLVLT-----HAELDLTR 45
 DB 83 LSEKDKRILITGGAGFVGSHTDLKMDGHEVTVDNFFTGKRNVEHIGHENFELIN 142

QY 46 QADVESFSEKPVVYVILAAKVGGIHANNTPADFIGVNLQ-IQTNVIHSAYEHGVKK- 103
 DB 143 HDVVEPLYEVQIYHLASPAS-----PPNYMYPKIKTKNTIGTLMNGLAKR 192

QY 104 ----LFLGSSCIYKPAQPIPESSALLTASLEP--TNEWYAIAGIAGIKTCOAYRIQHG 157
 DB 193 VGARLLLASTSEVYGD--PEVHPQSEDYMGHVNPIGPRACYDEGKRVATCMCYVMKQEG 250

QY 158 WDAISGMTNLYGP-----NDNFHPENSHVLPALMRPFHEAKVNGABEVVWGTGSPLEP 213
 DB 251 VEVVRVARIFNTFGPRMHMDG-RVVSNFILQALQ-----GEPLTVYSGSQTRAF 299

QY 214 LHVDLADACVFLLDYRSGLEHVNIGSGQEVTTIRELAELVKEVVGEGKLGW----- 265
 DB 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTILEFAQLIKNLVSGSGEIQFLSEAQDDP 358

QY 266 DCTKPDGTPTKLMDSKSLASLGWTPKVSRLDGLSQTVDWYLNKV 309
 DB 359 QKRKPDIKKAKLM-----LGWEPVVPLEEGINKAIHYFRKEL 395

RESULT 14
 US-10-176-921-260
 ; Sequence 260, Application US/10176921
 ; Publication No. US2003002726A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: DeForge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330R1C288
 ; CURRENT APPLICATION NUMBER: US/10/176,921
 ; CURRENT FILING DATE: 2002-06-20
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 260
 ; LENGTH: 420
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-176-921-260

Query Match 13.1%; Score 215; DB 9; Length 420;
 Best Local Similarity 25.3%; Pred. No. 6.6e-13;
 Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

QY 1 MSDKSAK-IFVAGHRLGVSIAVRKLOEQGFTNLVLT-----HAELDLTR 45
 DB 83 LSEKDKRILITGGAGFVGSHTDLKMDGHEVTVDNFFTGKRNVEHIGHENFELIN 142

QY 46 QADVESFSEKPVVYVILAAKVGGIHANNTPADFIGVNLQ-IQTNVIHSAYEHGVKK- 103
 DB 143 HDVVEPLYEVQIYHLASPAS-----PPNYMYPKIKTKNTIGTLMNGLAKR 192

QY 104 ----LFLGSSCIYKPAQPIPESSALLTASLEP--TNEWYAIAGIAGIKTCOAYRIQHG 157
 DB 193 VGARLLLASTSEVYGD--PEVHPQSEDYMGHVNPIGPRACYDEGKRVATCMCYVMKQEG 250

QY 158 WDAISGMTNLYGP-----NDNFHPENSHVLPALMRPFHEAKVNGABEVVWGTGSPLEP 213
 DB 251 VEVVRVARIFNTFGPRMHMDG-RVVSNFILQALQ-----GEPLTVYSGSQTRAF 299

QY 214 LHVDLADACVFLLDYRSGLEHVNIGSGQEVTTIRELAELVKEVVGEGKLGW----- 265
 DB 300 QYVSDLVNGLVALMNS-NVSSPNLGNPEHTILEFAQLIKNLVSGSGEIQFLSEAQDDP 358

QY 266 DCTKPDGTPTKLMDSKSLASLGWTPKVSRLDGLSQTVDWYLNKV 309
 DB 359 QKRKPDIKKAKLM-----LGWEPVVPLEEGINKAIHYFRKEL 395

RESULT 15
 US-10-227-884-240
 ; Sequence 240, Application US/10227884
 ; Publication No. US20030027988A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Gerritsen, Mary
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Grimaldi, J. Christopher
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stephan, Jean-Philippe F.
 ; APPLICANT: Watanabe, Colin L.
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3530PIC79
 ; CURRENT APPLICATION NUMBER: US/10/227,884
 ; CURRENT FILING DATE: 2002-08-26
 ; PRIOR APPLICATION NUMBER: 10/119,480
 ; PRIOR FILING DATE: 2002-04-09
 ; PRIOR APPLICATION NUMBER: 60/059113
 ; PRIOR FILING DATE: 1997-09-17
 ; PRIOR APPLICATION NUMBER: 60/062287
 ; PRIOR FILING DATE: 1997-10-17
 ; PRIOR APPLICATION NUMBER: 60/063549
 ; PRIOR FILING DATE: 1997-10-28
 ; PRIOR APPLICATION NUMBER: 60/064103
 ; PRIOR FILING DATE: 1997-10-31
 ; PRIOR APPLICATION NUMBER: 60/069873
 ; PRIOR FILING DATE: 1997-12-17
 ; PRIOR APPLICATION NUMBER: 60/078910
 ; PRIOR FILING DATE: 1998-03-20
 ; PRIOR APPLICATION NUMBER: 60/079294
 ; PRIOR FILING DATE: 1998-03-25
 ; PRIOR APPLICATION NUMBER: 60/079656
 ; PRIOR FILING DATE: 1998-03-26
 ; PRIOR APPLICATION NUMBER: 60/079728
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: 60/081819
 ; PRIOR FILING DATE: 1998-04-15
 ; PRIOR APPLICATION NUMBER: 60/081955
 ; PRIOR FILING DATE: 1998-04-15
 ; PRIOR APPLICATION NUMBER: 60/082804
 ; PRIOR FILING DATE: 1998-04-22
 ; PRIOR APPLICATION NUMBER: 60/084441
 ; PRIOR FILING DATE: 1998-05-06
 ; PRIOR APPLICATION NUMBER: 60/085323
 ; PRIOR FILING DATE: 1998-05-13
 ; PRIOR APPLICATION NUMBER: 60/085579

7	PRIOR APPLICATION NUMBER: 60/106464
7	PRIOR FILING DATE: 1998-10-30
7	PRIOR APPLICATION NUMBER: 60/106905
7	PRIOR FILING DATE: 1998-11-03
7	PRIOR APPLICATION NUMBER: 60/108787
7	PRIOR FILING DATE: 1998-11-17
7	PRIOR APPLICATION NUMBER: 60/108801
7	PRIOR FILING DATE: 1998-11-17
7	PRIOR APPLICATION NUMBER: 60/108849
7	PRIOR FILING DATE: 1998-11-18
7	PRIOR APPLICATION NUMBER: 60/112422
7	PRIOR FILING DATE: 1998-12-15
7	PRIOR APPLICATION NUMBER: 60/113296
7	PRIOR FILING DATE: 1998-12-22
7	PRIOR APPLICATION NUMBER: 60/113605
7	PRIOR FILING DATE: 1998-12-23
7	PRIOR APPLICATION NUMBER: 60/113621
7	PRIOR FILING DATE: 1998-12-23
7	PRIOR APPLICATION NUMBER: 60/115559
7	PRIOR FILING DATE: 1999-01-12
7	PRIOR APPLICATION NUMBER: 60/115565
7	PRIOR FILING DATE: 1999-01-12
7	PRIOR APPLICATION NUMBER: 60/115733
7	PRIOR FILING DATE: 1999-01-12
7	PRIOR APPLICATION NUMBER: 60/119549
7	PRIOR FILING DATE: 1999-02-10
7	PRIOR APPLICATION NUMBER: 60/123618
7	PRIOR FILING DATE: 1999-03-10
7	PRIOR APPLICATION NUMBER: 60/125259
7	PRIOR FILING DATE: 1999-03-19
7	PRIOR APPLICATION NUMBER: 60/125775
7	PRIOR FILING DATE: 1999-03-23
7	PRIOR APPLICATION NUMBER: 60/126773
7	PRIOR FILING DATE: 1999-03-29
7	PRIOR APPLICATION NUMBER: 60/127987
7	PRIOR FILING DATE: 1999-04-05
7	PRIOR APPLICATION NUMBER: 60/130232
7	PRIOR FILING DATE: 1999-04-21
7	PRIOR APPLICATION NUMBER: 60/131022
7	PRIOR FILING DATE: 1999-04-26
7	PRIOR APPLICATION NUMBER: 60/131270
7	PRIOR FILING DATE: 1999-04-27
7	PRIOR APPLICATION NUMBER: 60/131291
7	PRIOR FILING DATE: 1999-04-27
7	PRIOR APPLICATION NUMBER: 60/131445
7	PRIOR FILING DATE: 1999-04-28
7	PRIOR APPLICATION NUMBER: 60/134287
7	PRIOR FILING DATE: 1999-05-14
7	PRIOR APPLICATION NUMBER: 60/140650
7	PRIOR FILING DATE: 1999-06-22
7	PRIOR APPLICATION NUMBER: 60/140723
7	PRIOR FILING DATE: 1999-06-22
7	PRIOR APPLICATION NUMBER: 60/141037
7	PRIOR FILING DATE: 1999-06-23
7	PRIOR APPLICATION NUMBER: 60/144758
7	PRIOR FILING DATE: 1999-07-20
7	PRIOR APPLICATION NUMBER: 60/145698
7	PRIOR FILING DATE: 1999-07-26
7	PRIOR APPLICATION NUMBER: 60/146222
7	PRIOR FILING DATE: 1999-07-28
7	PRIOR APPLICATION NUMBER: 60/146963
7	PRIOR FILING DATE: 1999-08-03
7	PRIOR APPLICATION NUMBER: 60/149320
7	PRIOR FILING DATE: 1999-08-17
7	PRIOR APPLICATION NUMBER: 60/149638
7	PRIOR FILING DATE: 1999-08-17
7	PRIOR APPLICATION NUMBER: 60/151733
7	PRIOR FILING DATE: 1999-08-31
7	PRIOR APPLICATION NUMBER: 60/164418
7	PRIOR FILING DATE: 1999-11-09
7	PRIOR APPLICATION NUMBER: 60/166361
7	PRIOR FILING DATE: 1999-11-16
7	PRIOR APPLICATION NUMBER: 60/169445

; PRIOR FILING DATE: 1999-12-07
 ; PRIOR APPLICATION NUMBER: 60/169495
 ; PRIOR FILING DATE: 1999-12-07
 ; PRIOR APPLICATION NUMBER: 60/169835

Query Match 13.1%; Score 215; DB 9; Length 420;
 Best Local Similarity 25.3%; Pred. No. 6.6e-13;
 Matches 87; Conservative 53; Mismatches 138; Indels 66; Gaps 13;

Qy	1	MSDKSAK-IPVAGHGLVGS	AIVRKLEQGGFTNLVLT	-----HAELDLTR	45
Db	83	LSEKRRRIITGAGVGS	HLTDKLMMDGHEVTVDN	FTGKRNVHEHWIGHEN	FELIN 142
Qy	46	QADVESFPQEKPVYVILA	AAKVGGIHANNITYPAD	FIGVNLQ-IOTNVIHSA	YEHGVKK- 103
Db	143	HDVVEPLYIEVDQIYH	LSPAS-----PPNYM	NPITKTKNTIGTLNML	GLAKR 192
Qy	104	-----LLFLGSSCTYP	KPAPQIPESALLTASLEP	--TNEWYAIKTAGIKTC	OAYRIQHG 157
Db	193	VGARLLLASTSEVYGD	--PEVHPQSEDYMGHVN	PIGPRACYDEGKRVAE	TMCYVMKQEG 250
Qy	158	WDAISGMPTNLYGP	----NDNFHPENSHVLP	PALMRPFHEAKVNGAE	EVVWVGTSPLREF 213
Db	251	VEVRVARIFNTGPRMMD	NG-RVSNFILQALQ	-----GEPLTVYGS	GSQTRAP 299
Qy	214	LHVDDLADACVFLLD	RYSGLEHVNI	SGOEVTIRELAELVKE	VVPEGKLGW----- 265
Db	300	QYVSDLVNGLVALMNS	-NVSSPVNLGNPEHT	ILEFAQLIKNLVGS	SGSEIQFLSEAQDDP 358
Qy	266	CTKPDGTPRKMDSSK	LASLGWTPKVSRLD	LSQTYDWYLNKV	309
Db	359	QKRPDIKKAKLM	-----LGWEPV	VPLEEGLNKAIHY	FRKEL 395

Search completed: June 3, 2003, 05:32:16
 Job time : 532 secs

Result No.	Query			Description	
	Score	Match	Length	ID	Description
1	857	91.6	1583	10	US-09-318-271-1
2	391	41.8	1490	10	US-09-734-569-147
3	275.2	29.4	566	9	US-09-894-844-49
C 4	269.8	28.8	14187	9	US-09-114-170-121
C 5	265	28.3	536165	9	US-09-933-964-1
6	262.2	28.0	966	10	US-09-318-271-3
7	213.6	22.8	701	10	US-09-734-569-15
8	145.8	15.6	933	10	US-09-962-805-2
9	101.6	10.9	1340	10	US-09-318-271-5
C 10	59.8	6.4	282	10	US-09-294-093B-2058
C 11	58.8	6.3	447	9	US-10-025-380-296
C 12	58.8	6.3	447	10	US-09-922-217-296
C 13	58.8	6.3	447	10	US-09-833-263-296
14	44.8	4.8	656	10	US-09-815-343-293
C 15	43.8	4.7	846	10	US-09-770-445-638
16	38.6	4.1	492	10	US-09-920-300A-1521
17	38.6	4.1	492	12	US-10-033-528-1521
18	38.4	4.1	1332	9	US-09-938-842A-190
C 19	38	4.1	413	9	US-10-184-644-156

DB 109 A T T G T C C G C A A G C T T C A G G A C A A G G T T C A C C A A T C T C G T T C T T A A A C A C A C G C C G A G 168

```

QY 121 CTTGATCTACTCGTCAAGCCGATGTTGAATCCTTCTTTTCTCAAGAGAACCGATTTAT 180
Db 169 CTTGATCTACTCGTCAAGCCGATGTTGAATCCTTCTTTTCTCAAGAGAACCGATTTAT 228
QY 181 GTAATCTTAGCAGCAGCTAAAGTTGGTGGTATTACGCTAAACACCTATCTCGTGTAT 240
Db 229 GTAATCTTAGCAGCAGCTAAAGTTGGTGGTATTACGCTAAACACCTATCTCGTGTAT 288
QY 241 TTCAATCGGTCTCAATCTCCAGATTACAGCAATGTTGATCCACCTCCTCATATGACGCGT 300
Db 289 TTCAATCGGTCTCAATCTCCAGATTACAGCAATGTTGATCCACCTCCTCATATGACGCGT 348
QY 301 GTGAAGAAGCTTCTTCTTCTTGGATCATCTGATTTACCTAAATTTGCTCTCAGCCA 360
Db 349 GTGAAGAAGCTTCTTCTTCTTGGATCATCTGATTTACCTAAATTTGCTCTCAGCCA 408
QY 361 ATTCTCGAGTCTGCTTTGTTAAAGCATCGCTTGAACCACTAATGAGTGGTATGCTATT 420
Db 409 ATTCTCGAGTCTGCTTTGTTAAAGCATCGCTTGAACCACTAATGAGTGGTATGCTATT 468
QY 421 GCTAAGATCGCTGGGATTAAAGCTTGTACGGCTTATAGGATTCAGACGCGATGGATGCA 480
Db 469 GCTAAGATCGCTGGGATTAAAGCTTGTACGGCTTATAGGATTCAGACGCGATGGATGCA 528
QY 481 ATCTGGCATGCTACTAATCTCTATGTCCTTAATGACAAATTTCCACCGGAGAAATCT 540
Db 529 ATCTGGCATGCTACTAATCTCTATGTCCTTAATGACAAATTTCCACCGGAG---TCT 585
QY 541 CATGTGCTTCTGCTTATAGGAGGTTCCAGCGCGGAAAGTGAA--TGGACCGAGGA 599
Db 586 CATGTGCTTCTGCTTATAGGAGGTTCCAGCGCGGAAAGTGAA--TGGACCGAGGA 645
QY 600 AGTTGTGCTGTGGGTACAGGTAGTCC---GTTGAGGAGTCTTGTGATGTTGATGAT 654
Db 646 AGTTGTGCTGTGGGTACAGGTAGTCC---GTTGAGGAGTCTTGTGATGTTGATGAT 705
QY 655 TTGGCTGATGCTTGTGTTTCTTGTGATGCG---ATACAGCGGGTTGGAGCATGTTAAC 711
Db 706 TTGGCTGATGCTTGTGTTTCTTGTGATGCG---ATACAGCGGGTTGGAGCATGTTAAC 765
QY 712 ATTGGAAGTGGTCAAGAGTGAATATAGAGATTGGCTGAGTTGGTGAAGAGTTGTT 771
Db 766 ATTGGAAGTGGTCAAGAGTGAATATAGAGATTGGCTGAGTTGGTGAAGAGTTGTT 825
QY 772 GGTGTTGAAGGGAAGCTTGGATGGGATTGACCTAAGCCAGATGGCACCGGAGAACTT 831
Db 826 GGTGTTGAAGGGAAGCTTGGATGGGATTGACCTAAGCCAGATGGCACCGGAGAACTT 885
QY 832 ATGACAGCTCAAGCTCGCTCTTTGGGTGGACACCTAAGGTTTCTTATAGATGTT 891
Db 886 ATGACAGCTCAAGCTCGCTCTTTGGGTGGACACCTAAGGTTTCTTATAGATGTT 945
QY 892 CTGAGCCAACTTATGATGTTGTTGAAAGATGTTTGAACCGA 936
Db 946 CTGAGCCAACTTATGATGTTGTTGAAAGATGTTTGAACCGA 990

```

```

RESULT 2
US-09-734-569-147
; Sequence 147, Application US/09734569
; Patent No. US20020064816A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; APPLICANT: Reski, Ralf

```

```

; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; FILE REFERENCE: BASP-NAE-1332-99-US
; CURRENT APPLICATION NUMBER: US/09/734,569
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/171,101
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1
; SEQ ID NO 147
; LENGTH: 1490
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (347)..(1276)
; OTHER INFORMATION: c_pp001064043r
US-09-734-569-147

Query Match 41.8%; Score 391; DB 10; Length 1490;
Best Local Similarity 65.0%; Pred. No. 3,6e-119;
Matches 596; Conservative 0; Mismatches 315; Indels 6; Gaps 1;

QY 7 GACAAATCTGCCAAATCTTCGTCCGGGTCACTCGTGGTTCGTTGGTTCGATCTGCCATTGTC 66
Db 356 GACAAAGGACGCCAAGATCTTTGTTCTGGACACCGAGGTCTAGTAGGTGCGCTGTGTTT 415
QY 67 CGCAAGCTTCAGGAACAAGTTTCACCAATCTCGTTCCTTAAACACACGCCGAGCTTGTAT 126
Db 416 CGTGTCTTGAAGAAGGATGTTTAAACAAATTTGGTGATGAAGACTCATAAAGAATCTAGT 475
QY 127 CTCACCTCGTCAAGCGGATGTTGAATCTCTTTCTCAAGAGAACCCAGTTTATGTAATC 186
Db 476 CTTACCCGTACGACGAGG-----AATTTTCGACGCGGAAACCGATCTAGTCATC 529
QY 187 CTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTAAACACCACTATCTCTGCTGATTTCAAT 246
Db 530 CTAGCAGCTCGAGGTGGAGGCAATTCACGCAACACAGTACTTACCTCTGACAGATTCAAT 589
QY 247 GGTGTCAATCTCCAGATTTCAGACCAATGTGATCCACTTCGCATATGAGCAGCGTGTGAAG 306
Db 590 GCGGTGAATCTGCAGATCCAAACGAATGTCTGATGCTGTACAACTCTGGGGTGAAG 649
QY 307 AAGCTTCTCTCTTGGATCATCTGCTATTTACCCCTAAATTTGCTCTCAGCCAAATCTCT 366
Db 650 AAGCTTCTCTCTGCGGCTCTTCGTGTATCTACCCAAAGTTTGCACGATACCCATCGTT 709
QY 367 GAGTCTGTTGTTTAAACAGCATCGCTTGAACCACTAATAGTGGTATGCTATTCCTAAG 426
Db 710 GAGGAGTCGCTCTGACAGGCGCTTTGGAAGCTACAAACGAGTGGTATGCTGTAGCAAAG 769
QY 427 ATCGCTGGGATTAAGACTTGTTCAGGCTTATAGGATTCAGACGCGATGGATGCAATCTCT 486
Db 770 ATTGAGGAATCAAAATGTCAGGCTTACAGGCTGCAAGTATAATTTTCAGCGCCATTTCT 829
QY 487 GGCATGCTCTAATCTCTATGCTCTAATGACAAATTTCCACCGGAGAAATTTCTCATG 546
Db 830 GGAATGCCGACAAACCTCTACGGTCCCACGACAAATTTCCATCCCGAGAACTCCACGCTC 889
QY 547 CTTCTGCTCTTATAGGAGGTTTCCAGGCGGAAAGTGAATGAGCGGAGGAAAGTTGTG 606
Db 890 TTGCCAGCTTGTATCAGACGCTTTTTCAGGAGCTTAAGGTGAACGCGGCTTAAGGAAGTGT 949
QY 607 GTGTGGGATCAGGTAGTCCGTTGAGGAGTCTTTCGATGTTGATGATTTGGCTGATGCT 666
Db 950 GTGTGGGATCAGGTTCCTCCCATTCGTTGAGTTCCTTTCACGTTGAGGAGCTTGGCAGAG 1009
QY 667 TGTGTTTTCTTCTGGATCGATCAGCGGGTTGAGCATGTTAACTATGGAAGTGTCTCAA 726
Db 1010 ACAGTATTTCTGCTGAGAAATTTACTCCCGCATGAGCATGTCACATGCGCAGTGGCTCT 1069
QY 727 GAGTGAATATAGAGAGTTCGCTGAGTTGTGTTGAAAGAGTGTGTTTGAAGGGAAG 786

```

Db 1070 GAGGTCCTCAATCAGAGAACTCGCCGAAATGGTGAGAGAAAGTGGTTGGATTTCAGGGGAG 1129
QY 787 CTTGGATGGGATTCGACTAAGCCAGATGGCACACCGAGGAAATTTATGGACAGCTCAAAAG 846
Db 1130 CTGACATGGGATACTTCTTAAGCCTGATGGAATCTCCACGAAAGCTCATCGATAGCAGCAA 1189
QY 847 CTCGGCTCTTTGGTTGGACACCTAAGGTTTCTTTAGAGATGGTCTGAGCCAACTTAT 906
Db 1190 CTTGCCAAATGGGGTGGGAAGCGAGAAATCCCTCAAGGAAGGATTGGCAGAGACTTAC 1249
QY 907 GATTGGTATTTGAAGAA 923
Db 1250 AATGGTACTGTGAGAA 1266

RESULT 3

US-09-894-844-49
; Sequence 49, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; FILE REFERENCE: the M. tuberculosis Complex
; CURRENT APPLICATION NUMBER: US/09/894,844
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 966
; TYPE: DNA
; ORGANISM: M. tuberculosis
US-09-894-844-49

Query Match 29.4%; Score 275.2; DB 9; Length 966;
Best Local Similarity 56.2%; Pred. No. 9.7e-81;
Matches 517; Conservative 0; Mismatches 403; Indels 0; Gaps 0;

QY 6 TGACAAATCTGCCAAATCTTCGTGCGGGTCAATCGTGGTTGGTGGATCTGCCATTGT 65
Db 30 TGACCGCGGCGCGGGTCTACATCGCGGGCATCGCGCTGGTGGTGGTGGTGGTGGT 89
QY 66 CGCAAGCTTCAGGAACAAGGTTTCAACCAATCTCGTTCTTAAACAACACGCGGAGTTGA 125
Db 90 ACGCAGGTTTGGCGGCGGGGTTACCAACCTGCTGGTGGTGGTGGTGGTGGTGGTGG 149
QY 126 TCTCCTCTGTCAGCGGATGTTGAATCCTCTCTCTCTCAAGAGAACCGATTATGTAAT 185
Db 150 TCTGACGGATCGGGCGCGGAGTTCGACTTCGTTCTCGAGTCGAGCGCGGATCGTCAT 209
QY 186 CTTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTAAACAACCTATCTCTGCTGATTTCAT 245
Db 210 CGNCGGGGCGCGGGTGGCGGCTCTGGCCACGACACCTACCGCGCGGATTTCCT 269
QY 246 TGGTGTCAATCTCCAGATTACAGCAATGTGATCCACTCTGCAATATGACGCGGTGTGA 305
Db 270 GTCGGAACACCTCCAGATCCAGTCAACCTGCTGGATCGCGCGTGGCGCGGGTGGC 329
QY 306 GAAGCTTCTCTCTTGGATCATCTCGCATTTACCTTAATTTGCTCTCCAGCAATTC 365
Db 330 GCGGCTGCTGTTCTCGGCTCGTGGTGCATCTACCGAATCTGCGCGCGGCGGATCC 389
QY 366 TGAGTCTGCTTTGTTAAACAGCATCGCTTGAACCAACTAATAGTGGTATGCTATTCTAA 425
Db 390 GGAGAGCGGCTGCTCACCCTGCTGGAGCGGCAACAGCGCGGTACCGATCGCCAA 449

QY 426 GATCGCTGGGATTAAGACTTTGTCAGGCTTATAGGATTCAGACGCGATGGGATGCAATCTC 485
Db 450 AATCGCCGCGCATCTTGGCGTCCAGGCGGTGCGCGCCCAACATGCGCTCCGCTGGATCTC 509
QY 486 TGGCATGCTACTATCTATATGTCCTTAATCACAATTTCCACCGGAGAAATTTCTCATGT 545
Db 510 GCGGATGCCCAACACCTGTACGGGCCAGGCGCAACATTTTCGCCCTCGGCTCGCATCT 569
QY 546 GCTTCTGCTCTTATGAGGAGTTTCCAGGAGCGAAAGTGAATGGAGCGGAGAAATTTGT 605
Db 570 GCTCGCGGCACTCATCGCGCTATGACGAGGCCAAAGCCAGTGGCGCGCCCAACCTGAC 629
QY 606 GGTGTGGGTACAGGTAGTCCGTTGAGGGAGTTCTTGCAATGTTGATGTTGGTGTATGC 665
Db 630 CAACTGGGGCACCGGCGACCGCCCGAGGGAGTTGCTGCACTGCGACGACCTGGCGAGCG 689
QY 666 TTGTGTTTTCTTGTGATCGATACAGCGGGTTGGAGCATGTTAACTTGGAGTGGTCA 725
Db 690 ATGCTGTATCTGCTGGAAACATTTGACGCGGCCGACCCATGTCACGTTGGAAACCGGCAT 749
QY 726 AGAAGTGAATATTAGAGAGTTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTTTGAAGGAA 785
Db 750 CGACACACCATCGCGGAGATCGCGGAGATGTTGCGCTCGCGGTAGGCTATAGCGCGCA 809
QY 786 GCTTGGATGGGATTCGACTAAGCCAGATGGCAACCGAGGAAACTTTATGACAGCTCAA 845
Db 810 AACCCTGCTGGATCCAAAGCAACCGGACGGAACACCCACCAAACTGCTGGATGTTTGGT 869
QY 846 GCTCGGCTTTGGGTTGGACACCTAAGGTTTCTCTTACAGATGGTCTGAGCCAACTTA 905
Db 870 GCTACGGAGCGGGATGCGGCTTTCGATCGCGCTGCGGACGGCATCAGCGCGACGGT 929
QY 906 TGATTGGTATTTGAAGAATG 925
Db 930 GCGGTGATTCGCGAGCAG 949

RESULT 4

US-10-114-170-121/c
; Sequence 121, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386

```

REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 121:
SEQUENCE CHARACTERISTICS:
LENGTH: 14187
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 121:
US-10-114-170-121

```

Query Match	28.8%	Score 269.8;	DB 9;	Length 14187;
Best Local Similarity	57.1%	Pred. No. 3.5e-78;		
Matches 530;	Conservative 0;	Mismatches 372;	Indels 27;	Gaps 1;
Qy	22	ATCTTCGTCGGGGTCATCGTGGTTCGGATTCGCCATTGTCGCAAGCTTCAGAA	81	
Db	6374	ATTTTTATGCTGCTACCAAGGAATGGTTGGATCAGCTATATCCGACGGCTCAACAA	6315	
Qy	82	CAAGGTTTCACCAATCTGGTCTTTAAACACACCGGAGCTTGATCTCACTCGTCAAGCC	141	
Db	6314	CGTGATGATGTTGAGTTGGTTTTACGTACTCGGGATGAATGAACTGTTGGATAGTAGC	6255	
Qy	142	GATGTTTGAATCCTTCTTTCTCAGAGAGCCAGCTTTATGTAATCTTAGAGCAGCTAAA	201	
Db	6254	GCTGTTTGGATTTTTTTCTTCACAGAAAATCACCAGGTTTATTTGGCAGCAGCAAAA	6195	
Qy	202	GTTGGTGGTATTCACGGTTAAACACACCTATCCTGCTGATTTCAATGGTGTCCTCAAG	261	
Db	6194	GTCGGAGGTATTTTAGCTTAACAGTCTTATCTCTCCGATTTTATATATGCAATATAATG	6135	
Qy	262	ATTCAGACCAATGTGATCCACTCTGCATATGAGCACGGTGTGAAGAAGCTTCTCTTCCTT	321	
Db	6134	ATAGAGCGGAATGTCAATTCATGCTGCCACAAAAAATAATGTAAATAAATCTTTTCTCTC	6075	

Qy	322	GGATCATCTCGCATTTTACCTTAATTTTGGCTCTCTCAGCGCAATTCCTGAGTCGTGCTTGTGA	381
Db	6074	GGTTTCGTGCTGATATTTATCTCTAAGTAGCACACCAACCGATTTATGGAAGACGAATTAATTA	6015
Qy	382	ACAGCATCGCTTGAACCAACTAATAGAGTGGTATCTATTGCTAAGATCGCTGGGATTAAG	441
Db	6014	CNAGGGAAACTTTGAGCCAAACAAATGAACCTTATCTATCGCAAAAATTCGAGGTATTAAA	5955
Qy	442	ACTTGT CAGGCTTATAGATTTCACGACGATGGATGCAATCTCTGGCATGCCCTACTAAT	501
Db	5954	TTATGTGAATCTTTATAACCGTCAGTTTGGCGTGATTAACCGTTCAGTAATGCCCAACCAAT	5895
Qy	502	CTCTATGGTCTCTAATGACAAATTTCCACCCGGAGAAATTCCTCATGTGCTTCTGCTCTTATG	561
Db	5894	CTTTATGGTCCAAATGACAAATTTTTCATCAAGTAATTTCTCATGTGATTCGGCGCTTTTG	5835
Qy	562	AGGAGTTTCMAGAGCGAAAGTGAATGGAGCGGAGGAAGTTTGGTGTGGGGTACAGGT	621
Db	5834	CGCGCTTTTCATGATGCTGTGGAAAACAATTCCTCCGAATGTTGTGTGTTTGGGGAAGTGGT	5775
Qy	622	AGTCCGTTGAGGGAGTTCTTTCATGTTTGTATGATTTGGCTGATGCTGTGTGTTTTCTTCGCTG	681
Db	5774	ACTCCAAAGCGTGAATTTCTACATGTAGATGATATGGCTTCGCAAGCATTTATGTCAATG	5715
Qy	682	G-----ATCGATACAGCGGGTTGGAGCAATGTTTAAACATT	714
Db	5714	GAGATGCCATACGATATATGGCAAAAAATACTAAAGTAATGTTGTCTCATATCAATTAAT	5655
Qy	715	GGAAAGTGTCAAGAAGTGACTATTATAGAGAGTTGGCTCAGTTTGGTGAAGAGGTTGTGGT	774
Db	5654	GGACAGGTATTGACTGCACGATTTGTGTGAGCTTCGGGAAACAATAGCAAAAGTTGTAGGT	5595
Qy	775	TTTGAAGGGAAGCTTTGGATGGGATTGCATTAAGCCAGATGGCACACCGAGGAAACTTTATG	834
Db	5594	TATAAAGGGCATATTACGTTCGATACAAACAAAGCCGATGGAGCCCTCGAAAACACTACTT	5535

Qy	835	GACAGCTCAAA	CTCGGCTCTTTGGGTTGGACACCTTAAGGTTTCTCTTACAGATGCTCTG	894
Db	5534	GATGTACGGTCTCTT	CATCAACATAGGTTGGAAATCATATAAATTACCCCTTCAAGGGGTCCTT	5475
Qy	895	AGCAAACTTATGATTGGTATTTTGAAGAA	923	
Db	5474	GAATAATACATACAACATCGGTTTCTTTGAAA	5446	

```

RESULT 5
US-09-939-964-1/c
; Sequence 1, Application US/09939964
; Publication No. US20030054522A1
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; TITLE OF INVENTION: Plasmid
; FILE REFERENCE: CARP0068
; CURRENT APPLICATION NUMBER: US/09/939,964
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/214,808
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-939-964-1

```

Query Match	28.3%	Score 265,	DB 9,	Length 536165;
Best Local Similarity	56.6%	Pred. No. 1.5e-75;		
Matches 512;	Conservative 0;	Mismatches 390;	Indels 3;	Gaps 1;
Qy	22	ATCTTCGTCGCGGTCATCGTGGTTGGTTGGATCGCCATGTCGCGAAGCTTCAGGAA	81	
Db	6087	ATTTGGGTGCGAGGACACAAAGGATGGTCGGCAGCGCCATAATTCGATCGCTTGCTTC	6028	
Qy	82	CAAGGTTTCCACCAATCTCGTTCTTAAACACACGCGAGCTTGATCTCACTCGTCAAGCC	141	
Db	6027	GAGGATTGCGAGTCATCGTTGCAGATAGCA--AAAGCTTGATCTGACGCGCAAGAG	5971	
Qy	142	GATGTTGAATCCTCTTCTTCAAGAGAGCCAGTTTATGTAACTCTAGCAGCAGCTAAA	201	
Db	5970	GAACTTTCAGAAATTTCTATTAAAGGAAAGCCGACGCGTTCATAATGCGCGCAGGAAG	5911	
Qy	202	GTGTGGGTATTACACCTAACACACTATCTCGCTGATTTCAATGGTGTCAATCTCCAG	261	
Db	5910	GTGGGTGGGATCTCGCAAAATGATATATGCGCGCTGACTTCATCTATCAAAACCTTATC	5851	
Qy	262	ATTTCAGACCAATGTGATCCACTCTGATATGAGCAGGTGTGAAGAAGCTTCTCTTCCTT	321	
Db	5850	ATGAGGCTTAATGTCTATTGAGGCTCTCTCCGAGTGGCGTTGAAAAGCTCTCTTTCCCTT	5791	
Qy	322	GGATCATCTGCGAATTTACCTAAATTTGTCTCTCAGCCAAATTCCTGAGTGTGTTTGTTA	381	
Db	5790	GGATCGAGTTGCATATATCCGAAGTATGCGGGCAGCCCATGAAGGAAGAGGCTCTATT	5731	
Qy	382	ACAGCATCGCTTGAACCAACTAATGATGGTATGCTATTGCTAAGATCCCTGGATTAG	441	
Db	5730	ACCGGACCACTTGAAGCGGACCAACGAGTGGTATGCGATCGCCAAAATCGCCGCAATTAAG	5671	
Qy	442	ACTTGTTCAGGCTTATAGGATTTCAGCACGGATGGATGCAATCTCTGGCATGCCTACTAAT	501	
Db	5670	TTCTGTCAAGCGTATCGTAAAGCAATACGGCGCAACTTCATATCAGCCATGCCGACAAT	5611	
Qy	502	CTCTATGGTCTTAATGACAAATTTCCACCGGAGAAATTCATATGTGTCTCTGCTTTATG	561	
Db	5610	CTCTATGGGCCACGCGAATGAATTCGATCTTAATCTCAGCAGCAGTCTGCTGCTTAAATA	5551	

QY 562 AGAGGTTCCAGAGCGGAAAGTGAATGGAGCGGAGGAAGTTGTGTGGGTACAGGT 621
DB 5550 CGCAAGACACATGAGGCAAGATTAAGAGACCTTGGGTGCTTGTCTATATGGGGAAGCGC 5491
QY 622 AGTCGGTTGAGGAGTCTTGCATGTTGATGTTGGCTGATGCTTGTCTTCTTGTG 681
DB 5490 ACACCTACTCGAGACTTTTGTACAGTGAAGACTGCTCCGAGCCCTGCTTCTTCTTCT 5431
QY 682 GATCGATACAGCGGGTTGGAGCATGTTAAACATTTGGAAGTGGTCAAGAACTGACTATTAGA 741
DB 5430 AAGCATTTATCCGAACGGACACATTAACATAGGCTCCGGGGGGAATAGTATCAT 5371
QY 742 GAGTTGGCTAGTTGGTGAAGAGTGTGTTGTTTGAAGGGAAGCTCGATGGGATTCG 801
DB 5370 GAACTAGCCACATCTCTGCGCTGTTGTTGTTTAAAGGCGATATAGTCTTCGACACA 5311
QY 802 ACTTAGCCAGATGGCACCGGAAACCTTATGGACAGCTCAAGCTCCGCTTTTGGGT 861
DB 5310 TCAAGCCGAGGAAACGGACCGAAGCTTTTATAGCGAAGACTCTGTGCGATGGGT 5251
QY 862 TGGACACCTAAGGTTTCTCTTAGAGATGGTCTGAGCAAACTTATGATTGGTATTGAAG 921
DB 5250 TGGCGCCGAGACCTCGCTCGAGCTGGGACTGGCCAAATCTATGAATCGTTTGTACG 5191
QY 922 AATGT 926
DB 5190 AATGT 5186

RESULT 6

US-09-318-271-3
; Sequence 3, Application US/09318271A
; Patent No. US20020012979A1
; GENERAL INFORMATION:
; APPLICANT: Berry, Alan
; APPLICANT: Running, Jeffrey A.
; APPLICANT: Severson, David K.
; APPLICANT: Burlingame, Richard P.
; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"
; FILE REFERENCE: 3161-24
; CURRENT APPLICATION NUMBER: US/09/318,271A
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 60/125,073
; EARLIER FILING DATE: 1999-03-17
; EARLIER APPLICATION NUMBER: 60/125,054
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 60/088,549
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(966)
US-09-318-271-3

Query Match 28.0%; Score 262.2; DB 10; Length 966;
Best Local Similarity 56.5%; Pred. No. 2.1e-76;
Matches 526; Conservative 0; Mismatches 378; Indels 27; Gaps 1;
QY 21 AATCTTCGTCGGGTCATCGTGTGTTGGTTGGATCGCATGTCCGCAAGCTTCAGGA 80
DB 15 AGTTTTATGTGTGTCATCGCGGATGTGCGGTTCGGCATCAGCGCGCAGCTCGAACA 74
QY 81 ACAAGGTTTCCACCAATCTCGTTCTTAAACACACACGCGGAGCTTGATCTCACTCGTCAAGC 140
DB 75 GCGCGGTGATGTGGAAGTGTATTAGCGACCCGCGAGAGCTGAACCTGCTGGACGCGC 134
QY 141 CGATGTTGAATCCTTCTTCTCAAGAGAAGCGAGTTTATGTAATCTTAGCAGCAGCTAA 200

DB 135 CGCCGTGTCATGATTTCTTTGCGAGCGAAACGTATTGACCAAGGTCTATCTGCGCGCGCGCAA 194
QY 201 AGTTGGTGTATTTACCGTTAAACACACCTATCTCTGTGATTTTCAATTTGGTGTGCAATCTCCA 260
DB 195 AGTGGCGGCATTTGTCACACACCTATCCGGGGATTTTCACTACACAGAACATGAT 254
QY 261 GATTGAGACCAATGATCCACTCTGCATATAGCAGCAGTGTGGAAGAGCTTCTCTTCTCT 320
DB 255 GATTGAGAGCAACATCATTTACCGCGCATCAGAACGACGTTGAAACAAATGCTGCTTCTCT 314
QY 321 TGGATCATCTGTCATTTACCTTAAATTTGCTCTCTCAGCCAAATCTCTGAGTCTGCTTTGTT 380
DB 315 CGGATCGTCTGTCATCTACCCGAAATCGGCAAAACAGCCGATGGCAGAAAGCGAGTTGTT 374
QY 381 AACAGCATCGCTTGAACCAACTAATGAGTGGTATGCTATTGCTAAGATCGCTGGGATTA 440
DB 375 GCAGGGCAGCTGGAGCCGACTTACGAGCCTTATGCTATTGCCAAATCGCCGGGATCAA 434
QY 441 GACTTGTGAGGCTTATAGGATTCAGCAGGATGGGATGCAATCTCTGGCATGCTTACTAA 500
DB 435 ACTGTGCGAATCATACACCGCCAGTACGAGCGGATTAACCGCTCAGTCAATGCGGACCAA 494
QY 501 TCTCTATGTTCTATGACAAATTTCCACCCGAGAAATTTCTCATGTGCTTCTCTGCTTCTAT 560
DB 495 CCTGTACGGGCCACACGCAACTTCCACCCGAGTAAATTCGCATGTGATCCGACATTTGCT 554
QY 561 GAGGAGGTTCCACGAGGCGAAAGTGAATGGAGCGGAGGAAGTTGTGGTGTGGGTACAGG 620
DB 555 GCGTCGTTCCACGAGGCGACGCGACAGNATCGCCGAGCTGGTGGTATGGGCGAGCGG 614
QY 621 TAGTCCGTTGAGGAGTCTTTCGATGTTGATGATTTGGCTGATGCTGTGTTTCTTGTCT 680
DB 615 TACACCGATGCGCAATTTCTGACGTCGATGATGCGCGCGGCGAGCATTCATGTCAT 674
QY 681 GGATC-----GATACAGCGGTTGGAGCATGTTTACAT 713
DB 675 GGAGCTGGCGCATGAAGTCTGGCTGGAGAACACCCAGCGCATGTTGTCGACATTAACT 734
QY 714 TGGAAAGTGTCAAGAAAGTGAATTTAGAGAGTTGGCTGAGTTGGTGAAGAGAGTTCTGG 773
DB 735 CGGCGAGCGGCTTGACTGCATCTCCGCGAGCTGGCGCAACCATCGCCAAAGTGGTGG 794
QY 774 TTTTGAAGGGAAGCTTGGATGGGATTCGACTAAGCCAGATGGCACACCGAGGAACTTAT 833
DB 795 TTAACAAAGCGCGGTGTTTTTGTGTCAGCAACACCGATGGCAGCGCCGCAAACTGCT 854
QY 834 GGACAGCTCAAGAGCTCGGCTCTTTGGGTTGGACACCTAAGGTTTCTCTTAGAGATGCTCT 893
DB 855 GGATGTGACGCGCTCGCATCAGCTTGGCTGGTATCAGCAAAATCTCACTGGAAGCGGGCT 914
QY 894 GAGCCAAACTTATGATTTGTTTGAAGAAT 924
DB 915 TCCAGCACTTACCAGTGTGTTCTCTTGAGAAT 945

RESULT 7

US-09-734-569-15
; Sequence 15, Application US/09734569
; Patent No. US20020064816A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; APPLICANT: Reski, Ralf
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved

```
; TITLE OF INVENTION: in the synthesis of carbohydrates
; FILE REFERENCE: BASF-NAB-1332-99-US
; CURRENT APPLICATION NUMBER: US/09/734,569
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/171,101
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1
; SEQ ID NO 15
; LENGTH: 701
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(487)
; OTHER INFORMATION: 20_ppprot1_064_d07
US-09-734-569-15

Query Match      22.8%; Score 213.6; DB 10; Length 701;
Best Local Similarity 65.5%; Pred. No. 2.5e-60;
Matches 312; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 448 CAGGCTTATAGATTTCAGCAGGATGGATGCAATCTCTGGCATGCGCTACTACTATCTCTAT 507
Db      |||||
QY 508 GGTCTTAATGACAAATTCACCGGAGAAATTCATGCTCTTCCCTCTTATGAGGAGG 567
Db      |||||
QY 62  GGTCCCCACGACAAATTCATCCCGAGAACTCCACGCTTTCGCCAGCTTGTATCAGACGC 121
Db      |||||
QY 568 TTCCACGAGCGGAAATGGAATGAGCGGAGGAAGTTGTGGTGTGGGGTACAGGTAGTCGG 627
Db      |||||
QY 122 TTTCACGAGGCTAAGGTGAACGGCGCTAAGGAAGTGGTTGTGGGGATCAGGTTCCCA 181
Db      |||||
QY 628 TTGAGGAGGTTCTTGATGTTGATGATTTGGTGTGATGCTGTTGTTTCTTGTCTGATCGA 687
Db      |||||
QY 182 TTCCGTGAGTTCTTCCAGTGGACGACTTGGCAGAGGCAACAGTATTTCTGCTGCAGAA 241
Db      |||||
QY 688 TACAGCGGTTGAGCATGTTAATGATGGAAGTGTGCAAGAGTGAAGTATGATGAGAGTTG 747
Db      |||||
QY 242 TACTCCGCGATGAGCATGTCAATGCGGAGTGGCTCTGAGGCTCTCAATCAAGGAATC 301
Db      |||||
QY 748 GCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGGAAGCTTTGGATGGGATTCGCAATAG 807
Db      |||||
QY 302 GCGGAATGGTGAAGAGTGGTTGGATTTCAAGGGGCACTGACATGGGATCTTCTAAG 361
Db      |||||
QY 808 CCAGATGGCACACCGAGGAACTTATGGACAGCTCAAGCTCGCGTCTTGGGTTGGACA 867
Db      |||||
QY 362 CCTGATGGAACCTCCACGAAAGCTCATCGATAGCAGCAAACTTGCACCAATGGGGTGGCA 421
Db      |||||
QY 868 CCTAAGGTTTCTCTTAGAGATGCTCTGAGCCAACTTATGATGTTGTTTGAAGAA 923
Db      |||||
QY 422 GCGAAGATTTCCCTCAAGGAAGATTGGCAGAGACTTACAAATGGTACTGTGAGAA 477
Db      |||||

RESULT 8
US-09-962-805-2
; Sequence 2. Application US/09962805
; Patent No. US20020058313A1
; GENERAL INFORMATION:
; APPLICANT: RENKONEN, Risto
; APPLICANT: MATILA, Pirkko
; APPLICANT: HIRVAS, Laura
; APPLICANT: HORTLING, Solveig
; APPLICANT: KALLIOINEN, Tuula
; APPLICANT: KAURANEN, Sirka-Liisa
; APPLICANT: JAERVINEN, Nina
; APPLICANT: MAEKI, Minna
; APPLICANT: NIITYMAEKI, Jaana
; APPLICANT: RAEBINAE, Jarkko
; TITLE OF INVENTION: USE OF RECOMBINANT ENZYMES FOR PREPARING GDP-L-FUCOSE AND FUCOSYL
; TITLE OF INVENTION: GLYCANS
; FILE REFERENCE: 2242/50463
```

```
; CURRENT APPLICATION NUMBER: US/09/962,805
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FI 20002114
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-09-962-805-2

Query Match      15.6%; Score 145.8; DB 10; Length 933;
Best Local Similarity 50.4%; Pred. No. 1.1e-37;
Matches 411; Conservative 0; Mismatches 397; Indels 7; Gaps 2;

QY 117 CGAGCTTGATCTCACTCOTCAAGCGGATGGTGAATCTTCTTCTCTCAAGAGAAGCCAGT 176
Db      |||||
QY 111 CGAATTGTATTTTGTGGATAAAGACACAGTTTCAAGCCTATTTTGAAGAATAACAGCCTAC 170
Db      |||||
QY 177 TTATGTAATCCTAGCAGCAGCTAAAGTTGGTGGTATTTCACGCTAACACACCTATCTCTGC 236
Db      |||||
QY 171 AGGCATATCATTTGTGCCGGGAGATGGGGGCAATTGTGCAAAACATGAAGATCTTTTC 230
Db      |||||
QY 237 TGATTTCAATGGTGTCAATCTCCAGATTTCAGACCAATGTGATCCACTCTGCATATGAGCA 296
Db      |||||
QY 231 AACTTACATGTTGAGAAATTTACTCATGGGTTTGTATCTTTTCTAGCGCTTTAGATTT 290
Db      |||||
QY 297 CGGTGTGAAGAGCTTCTTCTTCTTGGATCATCTCTGATATTTACCTTAATTTGCTCTCA 356
Db      |||||
QY 291 GGGCGTGAAGAAAGCATTAAATCTAGCGAGCTTCTTGGCCTTATCTTAATACGCCCTAA 350
Db      |||||
QY 357 GCCAATCTCTCAGTCTGCTTTGTTAAACAGCATCGTTGAACCACTAATGATGATGCTATGC 416
Db      |||||
QY 351 CCTTTAAAGAGAGCGAATTTTATGAACGGCTCTTTAGAACCAAGCAATGAAGCTACGC 410
Db      |||||
QY 417 TATTGCTAAGATCGCTGGGATTAAGACTTGTCAAGCTTTATAGGATTCACGACGATGGGA 476
Db      |||||
QY 411 TTTAGCCAAACTCTCTGTGATGAAGTATTCGTAATACGTGAGCGCTGAAAAGCGGTTTT 470
Db      |||||
QY 477 TGCATCTCTGGCATGCCCTACTAATCTCTATGCTCTTAATGACAAATTTCCACCCGGAGAA 536
Db      |||||
QY 471 TTATAAAACTCTAGTGGCTTGTAACTTTATGGCGAGTTTGACAAAGTTTGAAGAAAAGAT 530
Db      |||||
QY 537 TTCTCATGCTCTCTCTCTTATGAGGAGGTTCCACGAGCGAAAGTGAATGGAGCGGA 596
Db      |||||
QY 531 AGCGCAGATGATACAGGGCTTATTTGTAGATGACACCCCTAAATTTAAATAATGAAAA 590
Db      |||||
QY 597 GGAAGTTGTGTGGGTACAGGTAGTCTCGTTGAGGAGGTTCTTGGCATGTTGATGATTT 656
Db      |||||
QY 591 AAATTTTGGCATGTTGGGCGATGGCAGCGCCAGAAAGAGAGTATCTAAACGCTAAAGATTT 650
Db      |||||
QY 657 GGC-----TGATGCTTGTGTTTCTTCTGCTGATGATACAGCGGGTTGGAGCATGTTAAC 711
Db      |||||
QY 651 AGCCAGATTCTCGCTCTCGCTTATGAGAAATATCGCTCAAAATGCTAGCTGATG--AAT 708
Db      |||||
QY 712 ATTGGAAGTGTCAAGAAGTGAATTTAGAGAGTTGGCTGAGTTGGTGAAGAGAGTTCTTT 771
Db      |||||
QY 709 GTCGCTCTGGAGTGGATTAACAGCATTTGAAGATTTTACGAAAAGTCCGCTCAGGTTTTA 768
Db      |||||
QY 772 GGTTTTGAAGGGAAGCTTTGGATGGGATTTGCACTAAGCCAGATGGCACACCGAGGAACTT 831
Db      |||||
QY 769 GACTATAAGGGCGTGTGTTGTGAAGAATTCATCAAAACCGAGTGGCATGCAACAAAAGCTT 828
Db      |||||
QY 832 ATGGACAGCTCAAGAGCTCGCGCTCTTGGGTTGGACACCTTAAGGTTTCTCTTAGAGATGCT 891
Db      |||||
QY 829 ATGGATATTTCCAAAACAAAAGGCTTTAAATGGGAATTTAGAAATCCCTTTTAGACAGGCG 888
Db      |||||
QY 892 CTGAGCCAAACTTATGATGTTGTTTGAAGAAATGT 926
Db      |||||
QY 889 ATCAAGAGCTTAAGATGATTTTGAAGCTTTT 923
Db      |||||
```

RESULT 9

US-09-318-271-5
; Sequence 5, Application US/09318271A
; Patent No. US20020012979A1
; GENERAL INFORMATION:
; APPLICANT: Berry, Alan
; APPLICANT: Running, Jeffrey A.
; APPLICANT: Severson, David K.
; APPLICANT: Burlingame, Richard P.
; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"
; FILE REFERENCE: 3161-24
; CURRENT APPLICATION NUMBER: US/09/318,271A
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 60/125,073
; EARLIER FILING DATE: 1999-03-17
; EARLIER APPLICATION NUMBER: 60/125,054
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 60/088,549
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5:
; TYPE: DNA
; LENGTH: 1340
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(1040)
US-09-318-271-5

Query Match 10.9%; Score 101.6; DB 10; Length 1340;
Best Local Similarity 48.7%; Pred. No. 7.4e-23;
Matches 340; Conservative 0; Mismatches 349; Indels 9; Gaps 2;

QY 124 GATCTCACTCGTCAAGCCGATGTGAATCCTCTCTTTCTCAAGAGAGCCAGTTTATGTA 183
DB 213 GATCTCAGGATACAGACAGACCCGCGCTGTGTGAGAGGTCCACCCACACAGTC 272

QY 184 ATCTAGACAGAGCTAAAGTTGGTGTATTCAGCTTAACACACCTATCTCTGATTTTC 243
DB 273 ATCCATCTTGTGCAATGTGGGGGCTGTTCGGAATATCAATACATTTGGACTTC 332

QY 244 ATTGGTGCATCTCAGATTCAGACCAATGTATCCATCTGCATATAGACACGGTGTG 303
DB 333 TGGAGGAAAAAGTGCACATGAACACACGCTCTGCCTCGCCCTTGGAGTGGGGGCC 392

QY 304 AAGAGCTTCTCTCTTGGATCATCTGCAATTTACCTTAATTTGCTCTCAGCAATTT 363
DB 393 CGCAAGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452

QY 364 CTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
DB 453 GATGAGACCATGATCCACATGGGCTCTCCACACAGCAATTTGGGTACTCGTATGCC 512

QY 424 AGATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGGATGGGATGCAATC 483
DB 513 AAGAGGATGATCGAGCTGAGAACAGGGCTACTCTCCAGCAGTACGGCTGCACCTTCACC 572

QY 484 TCTGCGATCGCTTACTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
DB 573 GCTGTCATCCCCACCAAGCTTTTCGGGGCCCCACGACAACTTCAATCAGGATGGCCAC 632

QY 544 GTGCTTCTGCTGCTTATAGGAGGTTTCCACGAGGGGAAAGTGAATGGAGCGGAGGAGTT 603
DB 633 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 689

QY 604 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
DB 690 ACGGTGTGGGTACAGGGAATCCGCGGAGGCGATTCATATCTCGTGGACCTGGCCAC 749

QY 664 GCTTGTGTTTCTTCTGCTGATCATACAGCGGGTGGAGCATGTTA-----ACATGGA 717
DB 750 CTCCTTATCTGGGTCTCGGGAGTACAAATGAAGTGGAGGCCCATCTCTCTCCGTGGGC 809

QY 718 AGTGTCAAGAGTCACTATTAGAGAGTTGGTGTGAGTTGGTGAAGAGGTTGTGTGTTT 777
DB 810 GAGGAAGATGAGTCTCCATCAAGGAGCGAGCGGCTGTGTGAGCCATGAGCTTC 869

QY 778 GAAGGGAAGCTTGGATGGATGCTAAGCCAGATGG 815
DB 870 CATGGGAAGTCACCTTTGATACAAACCAAGTCGGATGG 907

RESULT 10
US-09-294-093B-2058/C
; Sequence 2058, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 2058
; LENGTH: 282
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700345468H1
; NAME/KEY: unsure
; LOCATION: 2, 6, 50
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-2058

Query Match 6.4%; Score 59.8; DB 10; Length 282;
Best Local Similarity 56.8%; Pred. No. 2.2e-09;
Matches 151; Conservative 0; Mismatches 108; Indels 7; Gaps 2;

QY 23 TCTTCGTGCGGGTCAATCGTGGTTGGTGGATCTGCCATTCGCGAAGCTTCAGGAAC 82
DB 279 TCTTCGTGCTCGGGCGCAACGATCTGTGCGTTAAACCATCTGTGTGCTGCTCTCATGC 220

QY 83 AAGTTTCCACCAATCTGTTCTTAAACACACGCGAGCTTCATCTCATCTGTCAGCG 142
DB 219 TTGTCTTCACTGCGTGTGCGCCACACCCAGCTTGACCTCACCCGCGAGGCTA 160

QY 143 ATGTTGAATCCTTCTTTTCTCAA-----GAGAAGCCAGTTTATGTAACTCTAGCAGCAG 196
DB 159 ACTTTGAGGCTTCTTTCACCGCGAGCGCGCGGCTATGTGTCTCATGGCCT 100

QY 197 CTAAAGTTGGTGGTATTCAGCTAACACCTATCTCTGCTGATTTCAAT-GGTGTCAAT 255
DB 99 CGAGTGTGGAAGCATCTCATGCCAACTCCGCTTCCCCACCGACTTCATNCACCTGACAAC 40

QY 256 CTCCAGATTCAGACCAATGTATCCA 281
DB 39 CTCTAGATCTAGATCAACTTTGTGCA 14

RESULT 11

US-10-025-380-296/C
; Sequence 296, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy

APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuqiu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasir A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C14
CURRENT APPLICATION NUMBER: US/10/025,380
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 296
LENGTH: 447
TYPE: DNA
ORGANISM: Homo sapiens
US-10-025-380-296

Query Match 6.3%; Score 58.8; DB 9; Length 447;
Best Local Similarity 46.4%; Pred. No. 6.4e-09;
Matches 192; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
QY 124 GATCTCACTCGTCAAGCCGATGTTGAATCTTCTTTCTCAAGAGAGCCAGTTTATGTA 183
Db 416 GATCTCAAGGATACAGACAGACCCGCGCCCTGTTGAGAGGTCCAACCTCACACGTC 357
QY 184 ATCTTAGCAGCAGTAAAGTTGGTATTTCAGCTTAACAACACCTATCTCTGATTTC 243
Db 356 ATCCATCTTGCTGCAATGTGGGGCCCTGTTCCGGAATATCAATACAAATTTGACTTC 297
QY 244 ATTGGTGCAATCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 303
Db 296 TGGAGGAAAAGCGTGCATGAACAGACACGCTCTGCACTCGCCCTTCGAGGTGGCGCC 237
QY 304 AAGAGCTCTCTCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATC 363
Db 236 CGCAAGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 177
QY 364 CTGAGTCTGCTGTTTAAACAGATCGCTTGAACCACTTAATGAGTGGTATGCTATTGCT 423
Db 176 GATGAGACCATGATCCAAATGGGCTCCCAACAGCAATTTGGGTACTCGTATGCC 117
QY 424 AAGATCGTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGCAGGATGGATGCAATC 483
Db 116 AAGAGGATGATCGAGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 57
QY 484 TCTGGCATGCTACTTAATCTCTATGCTCTATGATGATGATGATGATGATGATGATG 537
Db 56 GCTGTCTATCCCAACAGCTCTTCGGGGCCCAACAGCACTTCAACATCGAGGAT 3

RESULT 12
US-09-922-217-296/c
Sequence 296, Application US/09922217
Patent No. US20020076414A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yuqiu
APPLICANT: Smith, Carole Lynn
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun

APPLICANT: Clapper, Jonathan D.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.471C13
CURRENT APPLICATION NUMBER: US/09/922,217
CURRENT FILING DATE: 2001-08-03
NUMBER OF SEQ ID NOS: 1124
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 296
LENGTH: 447
TYPE: DNA
ORGANISM: Homo sapiens
US-09-922-217-296

Query Match 6.3%; Score 58.8; DB 10; Length 447;
Best Local Similarity 46.4%; Pred. No. 6.4e-09;
Matches 192; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
QY 124 GATCTCACTCGTCAAGCCGATGTTGAATCTTCTTTCTCAAGAGAGCCAGTTTATGTA 183
Db 416 GATCTCAAGGATACAGACAGACCCGCGCCCTGTTGAGAGGTCCAACCTCACACGTC 357
QY 184 ATCTTAGCAGCAGTAAAGTTGGTATTTCAGCTTAACAACACCTATCTCTGATTTC 243
Db 356 ATCCATCTTGCTGCAATGTGGGGCCCTGTTCCGGAATATCAATACAAATTTGACTTC 297
QY 244 ATTGGTGCAATCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 303
Db 296 TGGAGGAAAAGCGTGCATGAACAGACACGCTCTGCACTCGCCCTTCGAGGTGGCGCC 237
QY 304 AAGAGCTCTCTCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATC 363
Db 236 CGCAAGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 177
QY 364 CTGAGTCTGCTGTTTAAACAGATCGCTTGAACCACTTAATGAGTGGTATGCTATTGCT 423
Db 176 GATGAGACCATGATCCAAATGGGCTCCCAACAGCAATTTGGGTACTCGTATGCC 117
QY 424 AAGATCGTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGCAGGATGGATGCAATC 483
Db 116 AAGAGGATGATCGAGTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 57
QY 484 TCTGGCATGCTACTTAATCTCTATGCTCTATGATGATGATGATGATGATGATGATG 537
Db 56 GCTGTCTATCCCAACAGCTCTTCGGGGCCCAACAGCACTTCAACATCGAGGAT 3

RESULT 13
US-09-833-263-296/c
Sequence 296, Application US/09833263
Patent No. US20020110547A1
GENERAL INFORMATION:
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolk, John A.
APPLICANT: Meagher, Madeleine J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
FILE REFERENCE: 210121.471C12
CURRENT APPLICATION NUMBER: US/09/833,263
CURRENT FILING DATE: 2001-04-10
NUMBER OF SEQ ID NOS: 1093
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 296
LENGTH: 447
TYPE: DNA
ORGANISM: Homo sapien
US-09-833-263-296

Query Match 6.3%; Score 58.8; DB 10; Length 447;
Best Local Similarity 46.4%; Pred. No. 6.4e-09;
Matches 192; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

124	QY	GATCTCACTCGTCAAGCCGATGTTGAATCCTTCTTTTCTCAAGAGACCGAGTTATGTA	183
416	Db	GATCTCACGGATACAGCAGACGACCGCGCCCTGTTTGGAGAGGTCCAACATCACACACGTC	357
184	QY	ATCCTAGCAGCAGCTAAAGTTGGTGGTATTCA	243
356	Db	ATCCATCTTGTGTCGAATGGTGGGGGGCTGTTCCGGAATATCAATACAATTTGGACTTC	297
244	QY	ATTGGTGTCAATCTCCAGATTTCAGACCAATGTGATCCACTCTGCATATGACGACGGTGTG	303
296	Db	TGGAGGAAAAACGTGCACATGAACGACAAACGTCTGCACTCGGCTTCGAGGTGGCGCC	237
304	QY	AGAAGCTTCTCTTCCTGGATCATCTCTGCAATTTACCTTAAATTTGGTCTCTCAGCCAAAT	363
236	Db	CGCAAGGTGGTCTCCTGGCTGTCCACCTGTATCTTCCCTGCACAGACGACCTACCCGATA	177
364	QY	CTGTAGTCTGCTTTGTTAACAGCATCGCTTTGAACCAACTAATGAGTGGTATGCTATTGCT	423
176	Db	GATGAGACCATGATCCACAATGGGCTCCCCACAAACAGCAATTTTGGGTACTCGTATGCC	117
424	QY	AAGATCGCTGGGATTAAGACTTGTTCAGGCTTTATAGGATTCAGCACGGATGGGATGCAATC	483
116	Db	AGAGAGTATCGACGTGCAGAACAGGGCCTACTTCCAGCAGTACGGCTGCACCTTCACG	57
484	QY	TTGGCATCGCTACTAATCTCTATGGTCCCTAATGACAAATTTCCACCGGAGAT	537
56	Db	GCTGTATCCCCACCAACGTCTTTCGGGGCCCCACGACCACTTCAACATCGAGGAT	3

```

RESULT 14
US-09-815-343-253
; Sequence 253, Application US/09815343
; Patent NO. US2001005596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 253
; LENGTH: 656
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(656)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-343-253

```

Query Match	4.8%;	Score 44.8;	DB 10;	Length 656;
Best Local Similarity	46.7%;	Prod. No. 0.00037;		
Matches 142;	Conservative 0;	Mismatches 162;	Indels 0;	Gaps 0;
Qy	124	GATCTCACTCGTCAAGCCGATGTTGCAATCCTTTCTTTCTCAAGAGAACCAGTTTATGTA	183	
Db	144	GATCTCACGNTACAGACACACCGCGCCCTGTTTGAGAGGTCCACACACACGTC	203	
Qy	184	ATCTTAGCAGCAGCTAAAGTTGGTGGTATTCA CGCTAACAAACCTATCTCTGCTGATTTTC	243	
Db	204	ATCCATCTTGTGCAATGGTGGGGGGCCGTGTTCCGGAAATACAAATACAAATTTGACATTC	263	
Qy	244	ATTGGTGTCAATCTCCAGATTACAGACCAATGTGATCCACTCTGCATATGACACGGTGTG	303	
Db	264	TGGAGGAAAACGTGCACATGAAACGATCTGCACTCGGCCCTTTGAGTGGGGGCC	323	
Qy	304	AAGAAGCTTCTCTTCTTGGATCATCTTCGCAATTTACCTTAATTTGCTCTCCAGCCAATT	363	
Db	324	CGCAAGGTGGTGTCTGCTGCTGCACCTGTATCTTCCCTGACAAAGACGACCTACCCGATA	383	

Qy	364	CTGAGTCTGCTTTGTTTAA	CAGCATCGTCTGA	CCCACTAAT	CAGTGGTATGCTATTGCT	423
Db	384	GATGAGACCATGATCCAC	AATGGGCTCCCCA	CACAGCAATTTT	TGGTACTCGTATGCC	443
Qy	424	AAGA	427			
Db	444	AAGA	447			

```

RESULT 15
US-09-770-445-639/c
; Sequence 638, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Goralach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770,445
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 638
; LENGTH: 846
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-638

```

Query Match	4.7%	Score 43.8	DB 10	Length 846
Best Local Similarity	51.2%	Pred. No. 0.00095		
Matches 155	Conservative 0	Mismatches 142	Indels 6	Gaps 2
Qy	595	GAGGAAGTTCTGCTGCGGTACAGGTAGTCCGTTGAGGGAGTCTTGCATGTTGATGAT	654	
Db	603	GAGCCCATGCACTGTTTACCGTGATGGGAAAGCAGACAAGGAGTTTCCAAATGTTGTTCTTGAT	544	
Qy	655	TTGCGCTGATGCTGTGTTTTCTTGCTGGATTCGATACAGCGGGTTGGAGCATGTTAAACATT	714	
Db	543	CTGGTTGAAGGTTTGATGAGACTGATGGAAGGAGAAATG- - -TCGCCCCATCAACCTC	487	
Qy	715	GGAAGTGGTCAAGAAAGTGACTATTAGAGAGTTTCGCTCAGTTGCTGGAAGAGGTTGTTGGT	774	
Db	486	GGTAA CCGTGGTGAATTACGATGCTCGAGCTCGCTAAGTGTGTCACAGAGACAATTGAT	427	
Qy	775	TTTGAAGGGGAAGCTTTGGATGGGATGCACTTAAAGCCAGATGGCAACAACGAGGAAACTTTATG	834	
Db	426	CCGAATGCAAAACATAGAGTTCAGACCAAAACACAGAAAGACGACCCTCACAGAGAAAGCCT	367	
Qy	835	GACAGCTCAAGCTC- - -GGGTCCTTTGGTGTGACACCTTAAGGTTTCTCTTAGAGATGGT	891	
Db	366	GACATCACAAAGCCCAAGAGCTTTTAGTTTGGGAACCAAGAGTCTCTCTTCGTCAAGGA	307	

Oy 892 CTG 894
|||
Db 306 CTG 304

Search completed: June 3, 2003, 05:01:27
Job time : 194 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 04:45:36 ; Search time 65 Seconds
(without alignments)
989.026 Million cell updates/sec

Title: US-10-089-014-1
Perfect score: 1646
Sequence: 1 MSDKSAKIFVAGHGLVGS.....SLRDGLSQTYDWYLNKNCNR 312

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_muc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1646	100.0	312	10	O49213 arabidopsis
2	1478	89.8	328	10	Q91MU0 arabidopsis
3	971	59.0	339	2	O86293 mycobacteri
4	971	59.0	339	2	O86298 mycobacteri
5	969	58.9	321	2	Q9RCB2 yersinia ps
6	956.5	58.1	321	2	Q56873 yersinia en
7	955	58.0	322	16	P71791 mycobacteri
8	944.5	57.4	321	16	Q8XGS8 salmonella
9	944.5	57.4	321	16	Q8X4R4 escherichia
10	928.5	56.4	321	2	Q9F7A3 salmonella
11	915.5	55.6	321	16	Q9S5F8 escherichia
12	915.5	55.6	323	2	O85340 escherichia
13	895.5	54.4	331	16	Q8U6L7 agrobacteri
14	893	54.3	360	2	Q8VU14 bacteroides
15	882	53.6	317	12	Q84611 paramecium
16	877.5	53.3	309	16	Q97H34 clostridium

17	848	51.5	322	16	Q98AU4 rhizobium 1
18	773.5	47.0	332	2	Q9JN55 coxiella bu
19	686	41.7	346	16	Q9PMM9 campylobact
20	672	40.8	289	2	Q9AQ09 bradyrhizob
21	621	37.7	208	2	Q9Z612 escherichia
22	596.5	36.2	246	16	Q9XDD7 brucella me
23	593.5	36.1	307	2	Q46720 escherichia
24	584.5	35.5	307	2	Q93Q29 salmonella
25	567.5	34.5	314	16	Q8YMW0 anabaena sp
26	565.5	34.4	312	16	P72585 synechocyst
27	547.5	33.3	308	2	O87152 vibrio chol
28	547.5	33.3	308	2	O51840 vibrio chol
29	540.5	32.8	307	2	Q8VQ41 escherichia
30	538.5	32.7	307	2	Q9F119 escherichia
31	535.5	32.5	308	2	O34227 synechococ
32	531.5	32.3	221	2	Q9R6T5 synechococ
33	531	32.3	287	2	Q9ZAY6 anabaena sp
34	505	30.7	310	16	O24886 helicobacte
35	502	30.5	310	2	O84974 helicobacte
36	495	30.1	310	16	Q9ZN21 helicobacte
37	437.5	26.6	321	5	Q9W1X8 drosophila
38	408	24.8	306	10	Q9SMD7 laminaria d
39	365.5	22.2	315	5	Q21632 caenorhabdi
40	345.5	21.0	114	16	Q8YBP7 brucella me
41	313.5	19.0	325	2	Q9FB21 streptomyce
42	273.5	16.6	309	17	Q8TXF0 methanopyru
43	270.5	16.4	310	17	Q9UXJ4 sulfolobus
44	266	16.2	309	16	Q9WYX9 thermotoga
45	264.5	16.1	310	16	Q55412 synechocyst

ALIGNMENTS

RESULT 1

ID O49213 PRELIMINARY; PRT; 312 AA.
AC O49213;
DC 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE GDP-4-keto-6-deoxy-D-mannose-3, 5-epimerase-4-reductase
(Fragment).
DE GER1 OR T18K17.8 OR ATFX.
GN Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA ECOTYPE;
RA Bonin C.P., Potter I., Vanzin G.F., Reiter W.-D.;
RT "A bifunctional epimerase-reductase completes the de novo synthesis of
GDP-L-fucose in Arabidopsis.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome I BAC T18K17 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Nakayama K., Maeda Y., Wang X., Jigami Y.;
RT "Expression of the genes for GDP-fucose synthesis in yeast
Saccharomyces cerevisiae.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF045286; AAC02703.2;
DR EMBL; AC010556; AAG52124.1;
DR EMBL; AB034806; BAA95670.1;

DR HSP; P32055; 1BSV.
 FT NON TER 312 312
 SQ SEQUENCE 312 AA; 34477 MW; 4F5E77D9FA492C0B CRC64;
 Query Match 100.0%; Score 1646; DB 10; Length 312;
 Best Local Similarity 100.0%; Pred. No. 4.9e-138;
 Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGTNLVLKTHAELDLTROADVSEFSSQKPY 60
 DB 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGTNLVLKTHAELDLTROADVSEFSSQKPY 60
 QY 61 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAYEHGVKKLLFLGSSCIYPKFAPQ 120
 DB 61 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAYEHGVKKLLFLGSSCIYPKFAPQ 120
 QY 121 IPESALLTASLEPTNEWYAIKIAIKTCQAVRIQHGWDALSGMPTNLYGPNDFHPENS 180
 DB 121 IPESALLTASLEPTNEWYAIKIAIKTCQAVRIQHGWDALSGMPTNLYGPNDFHPENS 180
 QY 181 HVLPMALRRPHEAKVNGAEVVMGTGSPLEFLHVDLADACVFLLDYSGLEHVNIGS 240
 DB 181 HVLPMALRRPHEAKVNGAEVVMGTGSPLEFLHVDLADACVFLLDYSGLEHVNIGS 240
 QY 241 GOEVTIRELAELVKEVVGEGKLGWDCPKDGTTPRKLMDSSKLASLGWTPKVSRLRDLSQ 300
 DB 241 GOEVTIRELAELVKEVVGEGKLGWDCPKDGTTPRKLMDSSKLASLGWTPKVSRLRDLSQ 300
 QY 301 TYDWYLNKVCNR 312
 DB 301 TYDWYLNKVCNR 312

RESULT 2
 Q9LMUO PRELIMINARY; PRT; 328 AA.
 ID 09LMUO
 AC 09LMUO
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE F2H15.12 protein (putative GDP-L-fucose synthetase).
 GN F2H15.12 OR ATG167890.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLOMBIA;
 RA Sakano H., Liu S.X., Etgu P., Lee J.M., Lenz C., Pham P., Toriumi M.,
 RA Yu G., Chan A., Chung M., Goldsmith A., Liu A., Smith A., Vaysberg M.,
 RA Altafi H., Brooks S., Buehler E., Chao Q., Conn L., Conway A.B.,
 RA Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M.,
 RA Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
 RA Federspiel N.A., Theologis A.;
 RT "The sequence of BAC F2H15 from Arabidopsis thaliana chromosome 1.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayaishiraki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Koeseema E., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RT "Full length cDNA of gene At1g17890 (GI:15220878).";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC034106; AAF97269.1; -;
 DR EMBL; AY063880; AAL36236.1; -;
 DR HSP; P32055; 1BSV.

DR InterPro; IPR002198; ADH short.
 DR PROSITE; PS00061; ADH SHORT; UNKNOWN 1.
 SQ SEQUENCE 328 AA; 36227 MW; 31C4479E9C5BIAC6 CRC64;
 Query Match 89.8%; Score 1478; DB 10; Length 328;
 Best Local Similarity 88.5%; Pred. No. 4.5e-123;
 Matches 276; Conservative 19; Mismatches 17; Indels 0; Gaps 0;
 QY 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGTNLVLKTHAELDLTROADVSEFSSQKPY 60
 DB 15 MLEKSAKIFVAGHRLVGSIVRKLQEQGTNLVLKTHAELDLTROADVSEFSSQKPY 74
 QY 61 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAYEHGVKKLLFLGSSCIYPKFAPQ 120
 DB 75 VILAAKVGGIHANNYPADFIGVNLQIQTNVHSAYEHGVKKLLFLGSSCIYPKFAPQ 134
 QY 121 IPESALLTASLEPTNEWYAIKIAIKTCQAVRIQHGWDALSGMPTNLYGPNDFHPENS 180
 DB 135 IPESALLTASLEPTNEWYAIKIAIKTCQAVRIQHGWDALSGMPTNLYGPNDFHPENS 194
 QY 181 HVLPMALRRPHEAKVNGAEVVMGTGSPLEFLHVDLADACVFLLDYSGLEHVNIGS 240
 DB 195 HVLPMALRRPHEAKVNGAEVVMGTGSPLEFLHVDLADACVFLLDYSGLEHVNIGS 254
 QY 241 GOEVTIRELAELVKEVVGEGKLGWDCPKDGTTPRKLMDSSKLASLGWTPKVSRLRDLSQ 300
 DB 255 GVEVTIRELAELVKEVVGEGKLGWDCPKDGTTPRKLMDSSKLASLGWTPKVSRLRDLSQ 314
 QY 301 TYDWYLNKVCNR 312
 DB 315 TYEWYLVNVQK 326
 RESULT 3
 O86293 PRELIMINARY; PRT; 339 AA.
 ID O86293
 AC O86293
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE GS8B protein.
 GN GS8B.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=11770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bull T.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,
 RA Hermon-Taylor J.;
 RT "A low G C content element in Mycobacterium avium subsp.
 RT paratuberculosis and M. avium subsp. silvaticum with homologous genes
 RT in M. tuberculosis";
 RL Microbiology 144:3413-3423 (1997).
 DR EMBL; AJ223833; CAAL1576.1; -;
 DR HSP; P32055; 1BSV.
 SQ SEQUENCE 339 AA; 37188 MW; A18AEBB191435B9C CRC64;
 Query Match 59.0%; Score 971; DB 2; Length 339;
 Best Local Similarity 59.2%; Pred. No. 5.5e-78;
 Matches 181; Conservative 50; Mismatches 75; Indels 0; Gaps 0;
 QY 3 DKSARKIFVAGHRLVGSIVRKLQEQGTNLVLKTHAELDLTROADVSEFSSQKPYVVI 62
 DB 28 DRATPVYIAGHRLVGSIVRKLQEQGTNLVRSRDEIDLTDRAATDFVSETRPQVII 87
 QY 63 LAAKVGGIHANNYPADFIGVNLQIQTNVHSAYEHGVKKLLFLGSSCIYPKFAPQIP 122
 DB 88 DAAARVGGIMANNYPADFLSNRIQTNLDDAAVAVRPRLLFLGSSCIYPKYAPQPIH 147

```

QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDATISGMTNLYGPNDFHPENSHV 182
DB 148 ESALLTGPLEPTNDAYAIKAGILQVAVRRQYGLAWISAMPTNLYGPGDNFSPSGSHL 207
QY 183 LPALMRRFHEAKVNGAEVWVGTSPLREFLHVDDADACVFLLDYSGLEHVNIGSQ 242
DB 208 LPALIRRYEEAKAGGAEEVNTWGTGTPRRELLHVDDLASACFLLEHFDGPNHNVNVTGV 267
QY 243 EYVIRELAELVKEVVGFEKLGWDCCTKPDGTPRKLMDSSKSLASLGWTPKVSURDGLSOTY 302
DB 268 DHSISEIADMVATAVGYIGETRWDPKPDGTPRKLDDVSALRELGNRPRIALKDGDATV 327
QY 303 DWYLNK 308
DB 328 SWYRTN 333

RESULT 4
ID O86298 PRELIMINARY; PRT; 339 AA.
AC O86298;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
DE GSBB protein.
GN GSBB OR MERA.
OS Mycobacterium avium subsp. silvaticum, and
OS Mycobacterium avium.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=44282, 1764;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.avium subsp. silvaticum;
RA Tizard M., Bull T., Millar D., Doran T., Martin H., Ford J.,
RA Hermon-Taylor J.;
RA "A low G+C content element in Mycobacterium avium subsp.
RT paratuberculosis and M. avium subsp. silvaticum with homologous genes
RT in M. tuberculosis.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RV [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.avium; STRAIN=2151;
RA Eckstein T.M., Lambert M.L., Brennan P.J., Belisle J.T., Inamine J.M.;
RA "Identification of a gene cluster involved in glycopeptidolipid
RT biosynthesis and of a gene cluster encoding daunorubicin resistance in
RT two strains of Mycobacterium avium serovar 2.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ223832; CAAL1572.1; -
DR EMBL; AF143772; RAD4220.1; -
DR EMBL; AF125999; AAD20374.1; -
DR HSSP; P32055; IBSV.
SQ SEQUENCE 339 AA; 37160 MW; 701B483E2CD3E7AF CRC64;

Query Match 59.0%; Score 971; DB 2; Length 339;
Best Local Similarity 59.2%; Pred. No. 5.5e-78;
Matches 181; Conservative 50; Mismatches 75; Indels 0; Gaps 0;

QY 3 DKSAKIFVAGHRLVGSALVRKLEQGFNVLKTHAELDLTROADVSEFFSQEKPVYVI 62
DB 28 DRATPVYIAGHRLVGSALVRFEAGSFNVLVRSRDEIDLTDRAATFDVSETRPQVII 87
QY 63 LAAKVGGIHANNTPADPIGVNLOIQTNVHSAYEHGVKLLFLGSSCIYPKFAPQPI 122

```

```

DB 88 DAAKVGGMANNTPADFLSENLRIOQNLLDAAVAVRPRLFLFGSSCIYPKYAPQPIH 147
QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDATISGMTNLYGPNDFHPENSHV 182
DB 148 ESALLTGPLEPTNDAYAIKAGILQVAVRRQYGLAWISAMPTNLYGPGDNFSPSGSHL 207
QY 183 LPALMRRFHEAKVNGAEVWVGTSPLREFLHVDDADACVFLLDYSGLEHVNIGSQ 242
DB 208 LPALIRRYEEAKAGGAEEVNTWGTGTPRRELLHVDDLASACFLLEHFDGPNHNVNVTGV 267
QY 243 EYVIRELAELVKEVVGFEKLGWDCCTKPDGTPRKLMDSSKSLASLGWTPKVSURDGLSOTY 302
DB 268 DHSISEIADMVATAVGYIGETRWDPKPDGTPRKLDDVSALRELGNRPRIALKDGDATV 327
QY 303 DWYLNK 308
DB 328 SWYRTN 333

RESULT 5
ID Q9RCB2 PRELIMINARY; PRT; 321 AA.
AC Q9RCB2;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE GDP-L-fucose synthetase.
GN FCL.
OS Yersinia pseudotuberculosis (type O:1b).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=109458;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA3606;
RX MEDLINE=20392461; PubMed=10931327;
RA Skurnik M., Peippo A., Ervela E.;
RT "Characterization of the O-antigen gene clusters of Yersinia
RT pseudotuberculosis and the cryptic O-antigen gene cluster of Yersinia
RT pestis shows that the plague bacillus is most closely related to and
RT has evolved from Y. pseudotuberculosis serotype O:1b.";
RL Mol. Microbiol. 37:316-330(2000).
DR EMBL; AJ251712; CAB63301.1; -
DR HSSP; P32055; IBSV.
DR InterPro; IPR00534; Semialdh dh.
DR Pfam; PF01118; Semialdehyde dh; 1.
SQ SEQUENCE 321 AA; 36289 MW; 10F97CC0D2ABCCB5 CRC64;

Query Match 58.9%; Score 969; DB 2; Length 321;
Best Local Similarity 59.1%; Pred. No. 7.7e-78;
Matches 188; Conservative 48; Mismatches 72; Indels 10; Gaps 2;

QY 3 DKSAKIFVAGHRLVGSALVRKLEQGFNVLKTHAELDLTROADVSEFFSQEKPVYVI 62
DB 2 DKK-RVFVAGHRLVGSALVRQLENRNDIELIIRTELDLMSQSAVQKFFATEKIDELY 60
QY 63 LAAKVGGIHANNTPADPIGVNLOIQTNVHSAYEHGVKLLFLGSSCIYPKFAPQPI 122
DB 61 LAAKVGGIQANNTPAFPIYQNLMECNIIHAHLAQKLLFLGSSCIYPKLAQPM 120
QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIQHGWDATISGMTNLYGPNDFHPENSHV 182
DB 121 EBALLTGVLEPTNEPYAIKAGIKLCEYNRQYGRDYSRVNPTNLYGNDNFHPENSHV 180
QY 183 LPALMRRFHEAKVNGAEVWVGTSPLREFLHVDDADACVFLLD-----RYSGL 233
DB 181 IPALLRRFHEAKIRNDKEMVWVGTPKPMREFLHVDDMAAASVHVMELSDQIYQTNQPM 240
QY 234 EHVNTSGOEVITRELAELVKEVVGFEKLGWDCCTKPDGTPRKLMDSSKSLASLGWTPKVS 293
DB 241 SHINVTGVDCTIRELAETMAKVGVGTGNLVFDSTKPDGTPRKLMDVSRKLKWCYQIS 300
QY 294 LRDGLSOTYDWYLNKVCN 311

```



```

QY 183 LPALMRPFHEAKVNGAEVVVWGTSPLREFLHVDDADACVFLLDYRYSGLHVNIGSQ 242
DB 191 LPALIRYDEKASGAPNTNMGTTGPRRELLHVDDASACVYLLHFDGPTHVNVGTGI 250
QY 243 EVTIRELAELVKEVGFEGKLGWDCTKPDGTPKRLMDSSKLASLGWTPKVSRLDGLSQTY 302
DB 251 DHTIGIEAWASAVGSGTETRWDPKPDGTPKRLDVSRLVREAGWRPSIALRDGTEATV 310
QY 303 DWY 305
DB 311 AWY 313

RESULT 8
Q8XGS8 PRELIMINARY; PRT; 321 AA.
AC Q8XGS8;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DE Bifunctional GDP fucose synthetase in colanic acid biosynthesis
DE (EC 1.-) (GDP-fucose synthetase) (EC 5.1.1.3.-).
GN WCAG OR STM2108 OR STV2320.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Churcher C., Mungall K.L., Bentley K., Chillingworth T., Connerton P.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque R., Hien T.F., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852 (2001).
RL EMBL; AE008793; AAL21012.1; -
DR EMBL; AL627273; CAD02471.1; -
DR InterPro; IPR001509; Epimerase_Dh.
DR Pfam; PF01370; Epimerase; 1
KW Oxidoreductase; isomerase; Complete proteome.
SQ SEQUENCE 321 AA; 35815 MW; 2139C7B9C6815D4B CRC64;

Query Match 57.4%; Score 944.5; DB 16; Length 321;
Best Local Similarity 58.5%; Pred. No. 1.2e-75;
Matches 182; Conservative 53; Mismatches 67; Indels 9; Gaps 1;

QY 7 KIFVAGHRLVGSIAVRKIQEQGFTNLVLTKEALDLTRQADVSEFSEKPYVILAAA 66
DB 5 RIFVAGHGMVGSIAVRQLAQRGDVLRTRDELDDLDGRAVQAFAGAGIDQVYLA 64
QY 67 KVGGIHANNTPADFTGVNLIQTNTVHSAYEHGVKKLLFLGSSCIYPKFAQPIPE 126
DB 65 KVGGIHANNTPADFTGVNLIQTNTVHSAYEHGVKKLLFLGSSCIYPKFAQPIPE 124

```

```

QY 127 LTASLEPTNEWAIKIAIKTKQAYRIQHGWDALISGMTNLYGPNDFHPNSHVLPA 186
DB 125 LQGTLEPTNEPVAIAKIAIKLCESYNOYGRDYRSMPTNLYGPHDFPNDSHVIPA 184
QY 187 MRRFHEAKVNGAEVVVWGTSPLREFLHVDDADACVFLLDYRYSGLHVN 237
DB 185 LREFHEAQAQSHAPEVVVWGSGTMRREFLHVDDMAAASHIMELAREVWQENTAPMLSHIN 244
QY 238 IGSQGVITIRELAELVKEVGFEGKLGWDCTKPDGTPKRLMDSSKLASLGWTPKVSRLD 297
DB 245 VGTGVDCITIRELAQTIKVGQYGRVVFDAARPDGTPKRLDVSRLVREAGWRPSIALRD 304
QY 298 LSQTYDWYLNKN 308
DB 305 LAGTYQWFLEN 315

RESULT 9
Q8X4R4 PRELIMINARY; PRT; 321 AA.
AC Q8X4R4;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE Putative nucleotide di-P-sugar epimerase or dehydratase (GDP-fucose
DE synthetase chain A).
GN WCAG OR Z3216 OR ECS2857.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22 (2001).
RL EMBL; AE005431; AAG57112.1; -
DR EMBL; AP002559; BAB36280.1; -
KW Complete proteome.
SQ SEQUENCE 321 AA; 36183 MW; 9511F1DFBD3D8058 CRC64;

Query Match 57.4%; Score 944.5; DB 16; Length 321;
Best Local Similarity 57.9%; Pred. No. 1.2e-75;
Matches 180; Conservative 55; Mismatches 67; Indels 9; Gaps 1;

QY 7 KIFVAGHRLVGSIAVRKIQEQGFTNLVLTKEALDLTRQADVSEFSEKPYVILAAA 66
DB 5 RIFVAGHGMVGSIAVRQLAQRGDVLRTRDELDDLDGRAVQAFAGAGIDQVYLA 64
QY 67 KVGGIHANNTPADFTGVNLIQTNTVHSAYEHGVKKLLFLGSSCIYPKFAQPIPE 126
DB 65 KVGGIHANNTPADFTGVNLIQTNTVHSAYEHGVKKLLFLGSSCIYPKFAQPIPE 124
QY 127 LTASLEPTNEWAIKIAIKTKQAYRIQHGWDALISGMTNLYGPNDFHPNSHVLPA 186

```

```
Db 125 LGTLEPTNEPYAIAKIAGIKLCEYNQYGRDYRSVMPNTNLYGPHDNFHPNSHVIPAL 184
Qy 187 MRRFHEAKVNGAEVWVGTSPLREFLHVDDLDAC-----VFLLDRYSGLEHVN 237
Db 185 LRPFHEATAQAPDVVVWVGSGTMRREFLHVDDDMAAASIHVMELAEHVWLENTQPMLSHIN 244
Qy 238 IGSQGVTTIRELAELVKEVVGFGKLGWCDKTPRKLMDSKSLASLGWTPKVSRLRG 297
Db 245 VGTGVDTCTIRELAQTAKVVGQYGRVVDASKPDGTPRKLLDVTRLHLQHWHEISLEAG 304
Qy 298 LSQTYDWYLN 308
Db 305 LASTYQWFLN 315
```

RESULT 10

```
Q9F7A3 PRELIMINARY; PRT; 321 AA.
AC Q9F7A3;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
GN GDP-L-fucose synthetase.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=20461159; PubMed=11004393;
RA Stevenson G., Lan R., Reeves P.R.;
RT "The colanic acid gene cluster of salmonella enterica has a complex
RT history."
RL FEMS Microbiol. Lett. 191:11-16(2000).
DR EMBL; AF285084; AAC24814.1; -.
DR HSSP; P32055; LBSV.
SQ SEQUENCE 321 AA; 35785 MW; BCB971B1127779F9 CRC64;
```

```
Query Match 56.4%; Score 928.5; DB 2; Length 321;
Best Local Similarity 57.9%; Pred. No. 3.1e-74;
Matches 180; Conservative 53; Mismatches 69; Indels 9; Gaps 1;
Qy 7 KIFVAGHGLVGSATVRKLBQEGFTNLVLTKEALDLTRQADVESFFSQEKPVYVILAAA 66
Db 5 RIFVAGHGMVGSATVRKLBQEGFTNLVLTKEALDLTRQADVESFFSQEKPVYVILAAA 64
```

```
Qy 67 KUGGIHANNTPADFGVNLQIOTNVHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 126
Db 65 KUGGIHANNTPADFGVNLQIOTNVHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 124
Qy 127 LTASLEPTNEWYAIKAGIKTCQAYRIQHGWDALSGMPTNLYGPNDFHPNSHVLPAL 186
Db 125 LGTLEPTNEPYAIAKIAGIKLCEYNQYGRDYRSVMPNTNLYGPHDNFHPNSHVIPAL 184
Qy 187 MRRFHEAKVNGAEVWVGTSPLREFLHVDDLDACVFLDD-----RYSGLEHVN 237
Db 185 LRPFHEAAASHAPEVWVGSGTMRREFLHVDDDMAAASIHVMELAEHVWQENTAPMLSHIN 244
Qy 238 IGSQGVTTIRELAELVKEVVGFGKLGWCDKTPRKLMDSKSLASLGWTPKVSRLRG 297
Db 245 VGTGVDTCTIRELAQTAKVVGQYGRVVDAAKPDGTPRKLLDVTRLHLQHWHEISLEAG 304
Qy 298 LSQTYDWYLN 308
Db 305 LASTYQWFLN 315
```

RESULT 11

```
Q9S5F8 PRELIMINARY; PRT; 321 AA.
ID Q9S5F8
```

```
AC Q9S5F8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE GDP-L-fucose pathway enzyme (Fucose synthetase).
GN WBHF OR FCI OR Z3197 OR ECS2838.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=184;
RX MEDLINE=99240837; PubMed=102222209;
RA Shimizu T., Yamaaki S., Tsukamoto T., Takeda Y.;
RT "Analysis of the genes responsible for the O-antigen synthesis in
RT enterohaemorrhagic Escherichia coli O157."
RL Microb. Pathog. 26:235-247(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RN Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AB008676; BAA77731.1; -.
DR EMBL; AE005429; AAG57093.1; -.
DR EMBL; AP002559; BAB36261.1; -.
DR HSSP; P32055; LBSV.
KW Complete proteome.
SQ SEQUENCE 321 AA; 36268 MW; 8C9A5311115E33791 CRC64;
```

```
Query Match 55.6%; Score 915.5; DB 16; Length 321;
Best Local Similarity 57.2%; Pred. No. 4.4e-73;
Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;
Qy 7 KIFVAGHGLVGSATVRKLBQEGFTNLVLTKEALDLTRQADVESFFSQEKPVYVILAAA 66
Db 5 RIFVAGHGMVGSATVRKLBQEGFTNLVLTKEALDLTRQADVESFFSQEKPVYVILAAA 64
Qy 67 KUGGIHANNTPADFGVNLQIOTNVHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 126
Db 65 KUGGIHANNTPADFGVNLQIOTNVHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 124
Qy 127 LTASLEPTNEWYAIKAGIKTCQAYRIQHGWDALSGMPTNLYGPNDFHPNSHVLPAL 186
Db 125 LGTLEPTNEPYAIAKIAGIKLCEYNQYGRDYRSVMPNTNLYGPNDFHPNSHVLPAL 184
Qy 187 MRRFHEAKVNGAEVWVGTSPLREFLHVDDLDACVFLDD-----RYSGLEHVN 237
Db 185 LRPFDAVENNSPNVWVGSGTMRREFLHVDDDMAAASIHVMELAEHVWQENTAPMLSHIN 244
Qy 238 IGSQGVTTIRELAELVKEVVGFGKLGWCDKTPRKLMDSKSLASLGWTPKVSRLRG 297
Db 245 IGTGIDCTICEIAETIAKVGKYGHTFTDTPKDGAPRKLLDVTLLHLQHWNHKTLHKG 304
```

QY 298 LSQTYDWYLN 308
| : : : : :
Db 305 LENTYNWFLEN 315

RESULT 12

Q85340 PRELIMINARY; PRT; 323 AA.
ID O85340;
AC O85340;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Fucose synthetase Fcl.
GN FCL.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C664-1992;
RX MEDLINE=98339851; PubMed=9673232;
RA Wang L., Reeves P.R.;
RT "Organization of Escherichia coli O157 O antigen gene cluster and
RT identification of its specific genes.";
RL Infect. Immun. 66:3545-3551(1998).
DR EMBL; AF061251; AAC32346.1; -.
DR HSSP; P32055; 1BSV.
SQ SEQUENCE 323 AA; 36531 MW; 382A03B17CABD6E9 CRC64;

Query Match 55.6%; Score 915.5; DB 2; Length 323;
Best Local Similarity 57.2%; Pred. No. 4.4e-73;
Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;
QY 7 KIFVAGHRLGVSIAVRKLOEGFTNLVLKTHAELDLTRQADVESFSEKPPVYVILAAA 66
| : : : : :
Db 7 RFIAGHQMGVSIAITRLKQDDVELVTRDELNLDSAVLDFSSQKIDQVYVILAAA 66
| : : : : :
QY 67 KVGGIHANNYPADFTIGVNLQITNVHSAYEHGVKLLFLGSSCIIYKPFAPQIPESAL 126
| : : : : :
Db 67 KVGILANSSYPADFIYENIMIEANVIAHAKNNVKKLFLGSSCIIYKPLAQHPIWDEL 126
| : : : : :
QY 127 LTASLEPTNEWYAIKAGIKTCQAYRIOHGWDAISGMTNLYGPNDFHNSHVPAL 186
| : : : : :
Db 127 LQKLEPTNEPYAIAKIAKLSQFYISQYGLNCVSMPTNIYGLNDFPDQSSHVIPAM 186
| : : : : :
QY 187 MRPFHAKVNGAEVVVWGTGSPRLREFLHVDDLADACVFLDDR-YG-----GLEHVN 237
| : : : : :
Db 187 LRRFHDVAVENSPNVVWGTGSPRLREFLHVDDNASASIVYEMWPDYIWKNTKVMLSHIN 246
| : : : : :
QY 238 IGSQGEVTIRELAELVKEVVGEGKLGWDCPTKPGTPRKLMDSSKLASLGWTPKVSRLRDG 297
| : : : : :
Db 247 IGTGIDCTICELAEITAKVVGKGTFTDTPKDGAPKLLDVTLLHQLGWNHKITLHKG 306
| : : : : :
QY 298 LSQTYDWYLN 308
| : : : : :
Db 307 LENTYNWFLEN 317

RESULT 13

Q8U6L7 PRELIMINARY; PRT; 331 AA.
ID Q8U6L7;
AC Q8U6L7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE GDP-fucose synthetase.
GN FCL OR ATU4790 OR AGR L.185.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Krespan W., Perry M.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,
RA Qurolo B., Goldman B.S., Cao Y., Askenazi M., Iartchouk O., Epp A., Liu F.,
RA Houmlel K., Gordon J., Vaudin M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Wollam C., Allinger M., Gursion J., Lomo C., Sear C., Strub G.,
RA Flanagan C., Crowell C.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AF009407; AAL45584.1; -.
DR EMBL; AF008206; AAK88659.1; -.
KW Complete proteome.
SQ SEQUENCE 331 AA; 36726 MW; 4DB36D98CEE75380 CRC64;
Query Match 54.4%; Score 895.5; DB 16; Length 331;
Best Local Similarity 55.5%; Pred. No. 2.7e-71;
Matches 166; Conservative 59; Mismatches 73; Indels 1; Gaps 1;
QY 7 KIFVAGHRLGVSIAVRKLOEGFTNLVLKTHAELDLTRQADVESFSEKPPVYVILAAA 66
| : : : : :
Db 18 RVWVAGHTMGVSALVREREN-CEILKVSRELDLTRYETEQWMAARPOVIFVAAA 76
| : : : : :
QY 67 KVGGIHANNYPADFTIGVNLQITNVHSAYEHGVKLLFLGSSCIIYKPFAPQIPESAL 126
| : : : : :
Db 77 KVGIIAANAYPADFTLYTNLISMNIMKSAADIGVKKLWMGSSCIIYKPFAPQIPESAL 136
| : : : : :
QY 127 LTASLEPTNEWYAIKAGIKTCQAYRIOHGWDAISGMTNLYGPNDFHNSHVPAL 186
| : : : : :
Db 137 LTGPLEPTNEPYAIAKIAKLSQFYISQYGLNCVSMPTNIYGLNDFPDQSSHVIPAM 196
| : : : : :
QY 187 MRPFHAKVNGAEVVVWGTGSPRLREFLHVDDLADACVFLDDRYSGLHVNIGSGQRTV 246
| : : : : :
Db 197 IRRMEAKISGQNKIVLWGTGSPRLREFLHVDDLADACVFLDDRYSGLHVNIGSGREIS 256
| : : : : :
QY 247 RELAEIVKEVVGEGKLGWDCPTKPGTPRKLMDSSKLASLGWTPKVSRLDGLSQTVDWY 305
| : : : : :
Db 257 RNLHLIAGIVGEGQIVFTSKPDGAPKLLDCSRNALGNSTVELRYGIQDLYEW 315
| : : : : :
RESULT 14
Q8VU14 PRELIMINARY; PRT; 360 AA.
ID Q8VU14;
AC Q8VU14;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE GDP-4-keto-6-deoxy-D-mannose-3, 5-epimerase-4-reductase.
GN FCL.
OS Bacteroides fragilis.
OC Bacteria; CFb group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
OC Bacteroides.
OX NCBI_TaxID=817;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 9343;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 05:01:31 ; Search time 35 Seconds
(without alignments)
856.970 Million cell updates/sec

Title: US-10-089-014-1
Perfect score: 1646
Sequence: 1 MSDKSAKIFVAGHRLVLSA.....SLRDGLSQTYDWLKNVNCNR 312
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1646	100.0	312	2	F96758
2	1478	89.8	328	2	F2H15.12 protein -
3	955	58.0	322	1	probable epiA prot
4	944.5	57.4	321	2	GDP-fucose synthet
5	944.5	57.4	321	2	GDP-fucose synthet
6	944.5	57.4	321	2	GDP-fucose synthet
7	943.5	57.3	321	1	hypothetical 36.1
8	915.5	55.6	321	2	fucose synthetase
9	915.5	55.6	321	2	fucose synthetase
10	895.5	54.4	331	2	hypothetical 34.7K
11	895.5	54.4	331	2	GDP-fucose synthet
12	882	53.6	317	2	hypothetical prote
13	877.5	53.3	309	2	nucleoside-diphosp
14	686	41.7	346	2	probable fucose sy
15	596.5	36.2	246	2	gdp-4-dehydro-D-rh
16	567.5	34.5	314	2	gndp-glucose dehyd
17	565.5	34.4	312	1	hypothetical prote
18	547.5	33.3	308	2	modulation protein
19	505	30.7	310	1	modulation protein
20	495	30.1	310	2	probable sugar nuc
21	401	24.4	271	1	hypothetical prote
22	365.5	22.2	315	2	hypothetical prote
23	345.5	21.0	114	2	gdp-fucose synthet
24	270.5	16.4	310	2	UDP-fucose 4-epim
25	266	16.2	309	2	hypothetical prote
26	264.5	16.1	310	2	hypothetical prote
27	264	16.0	341	2	conserved hypothet
28	258	15.7	314	2	probable dtdp-gluc
29	253.5	15.4	321	2	nucleotide sugar e

30	253	15.4	307	2	D75143	udp-glucose 4-epim
31	250	15.2	305	2	D64326	UDPglucose 4-epime
32	246	14.9	319	2	D86934	probable sugar-nuc
33	245	14.9	306	1	A71183	probable UDP-gluc
34	243	14.8	334	2	T44339	hypothetical prote
35	240	14.6	334	2	S70889	nucleotide sugar e
36	236	14.3	725	2	H96989	FUSON, Nucleoside
37	231.5	14.1	346	2	C72353	hypothetical prote
38	231	14.0	344	2	D75486	probable UDP-gluc
39	230	14.0	308	2	C84072	UDP-glucose 4-epim
40	227.5	13.8	394	2	A75303	UDP-glucose 4-epim
41	226.5	13.8	343	2	E84113	nucleotide sugar e
42	226	13.7	317	2	F90233	dtdp-Glucose 4,6-d
43	224.5	13.6	321	2	G95920	probable epimerase
44	223.5	13.6	329	2	G82785	dtdp-Glucose 4-6-d
45	221.5	13.5	318	2	H71145	probable UDP-gluc

ALIGNMENTS

RESULT 1

F96758
hypothetical protein T18K17.8 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C;Accession: F96758
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, R.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lurose, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F96758
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-312 <STO>
A;Cross-references: GB:AE005173; NID:G6598858; PIDN:AAF18712.1; GSPDB:GN00141
C;Genetics:
A;Gene: T18K17.8
A;Map position: 1
C;Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 100.0%; Score 1646; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 6e-131;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSDKSAKIFVAGHRLVLSAIVRKLOEQGTNLVLKTHAELDLTRQADVESFSSQKPVY	60
Db	1	MSDKSAKIFVAGHRLVLSAIVRKLOEQGTNLVLKTHAELDLTRQADVESFSSQKPVY	60
QY	61	VILAAKVGIGIHANNTPADFGVNLQIQTNIHSAYEHGVKLLFLGSSCIYKFAPOP	120
Db	61	VILAAKVGIGIHANNTPADFGVNLQIQTNIHSAYEHGVKLLFLGSSCIYKFAPOP	120
QY	121	IPESALLTASLEPTNEWYAIKAGIKTCQAVRIQHGWDIAISGMPTNLGPNDFNPENS	180
Db	121	IPESALLTASLEPTNEWYAIKAGIKTCQAVRIQHGWDIAISGMPTNLGPNDFNPENS	180
QY	181	HVLPALMRPFHAKNGAEVVMWGTGSLRFLHVDDLDADACVFLDDRYSGLEHVNIGS	240
Db	181	HVLPALMRPFHAKNGAEVVMWGTGSLRFLHVDDLDADACVFLDDRYSGLEHVNIGS	240
QY	241	GOEVITRELAEVLKVEVVGEGKLGWDCTKPDGTPRKLMDSSKLASLGWTPKVSRLRGLSQ	300
Db	241	GOEVITRELAEVLKVEVVGEGKLGWDCTKPDGTPRKLMDSSKLASLGWTPKVSRLRGLSQ	300
QY	301	TYDWLKNVNCNR	312

Db 301 TYDWLKNVNCNR 312
|||||

RESULT 2
B86314
F2H15.12 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 14-Dec-2001
C:Accession: B86314
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B86314
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-328 <STO>
A:Cross-references: GB:AE005172; NID:g9665067; PIDN:AAF97269.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 89.8%; Score 1478; DB 2; Length 328;
Best Local Similarity 88.5%; Pred. No. 9e-117;
Matches 276; Conservative 19; Mismatches 17; Indels 0; Gaps 0;

QY 1 MSKSAKIFVAGHRLVGSIAVRKLOEQGFTNLVLTHTAELDLTROADVSEFFSQKPVY 60
DB 15 MLEKSAKIFVAGHRLVGSIAVRKLOEQGFTNLVLTHTAELDLTROADVSEFFSQKPVY 74
QY 61 VILAAKVGGIHANNTPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYKPAPOP 120
DB 75 VILAAKVGGIHANNTPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYKPAPOP 134
QY 121 IPESALLTASLEPTNEWYAIKAGIKTCQAYRIHQHWDALSGMPTNLGPNDFHPENS 180
DB 135 IPESALLTASLEPTNEWYAIKAGIKTCQAYRIHQHWDALSGMPTNLGPNDFHPENS 194
QY 181 HVLPMALRRFHEAKNGAEVVGWGTSPLEFLHVDLADACVFLLDRIYSGLEHVNIGS 240
DB 195 HVLPMALRRFHEAKNGAEVVGWGTSPLEFLHVDLADACVFLLDRIYSGLEHVNIGS 254
QY 241 GOEVTIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKYSRLRDLGQ 300
DB 255 GVEVTIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKYSRLRDLGQ 314
QY 301 TYDWLKNVNCNR 312
DB 315 TYEWLENVQVK 326

RESULT 3
C70714
probable epiA protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: C70714
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70714

A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-322 <COL>
A:Cross-references: GB:Z79701; GB:AL123456; NID:g3261635; PIDN:CAB02026.1; PID:g1524259
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: epiA
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 58.0%; Score 955; DB 1; Length 322;
Best Local Similarity 59.7%; Pred. No. 9.6e-73;
Matches 181; Conservative 44; Mismatches 78; Indels 0; Gaps 0;

QY 3 DKSAKIFVAGHRLVGSIAVRKLOEQGFTNLVLTHTAELDLTROADVSEFFSQKPVY 62
DB 11 DRAARVYIAGHRLVGSIAVRKLOEQGFTNLVLTHTAELDLTROADVSEFFSQKPVY 70
QY 63 LAAAKVGGIHANNTPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYKPAPOP 122
DB 71 DAAARVGGILANDTYPADFLSENQIQVNLDAVAARVPRLLFLGSSCIYKPAPOP 130
QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIHQHWDALSGMPTNLGPNDFHPENS 182
DB 131 ESALLTASLEPTNEWYAIKAGIKTCQAYRIHQHWDALSGMPTNLGPNDFHPENS 190
QY 183 LPALMRFRHEAKNGAEVVGWGTSPLEFLHVDLADACVFLLDRIYSGLEHVNIGSQ 242
DB 191 LPALMRFRDEAKNGAEVVGWGTSPLEFLHVDLADACVFLLDRIYSGLEHVNIGSQ 250
QY 243 EYVIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKYSRLRDLGQ 302
DB 251 DHTIGEIAEMVASAVGSGTETWDPSKPDGTPRKLMDSSKLASLGWTPKYSRLRDLGQ 310
QY 303 DWY 305
DB 311 AWY 313

RESULT 4
A10768
GDP-fucose synthetase (EC 5.1.3.-) [imported] - Salmonella enterica subsp. enterica sero-
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: A10768
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Party, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero-
A:Reference number: AB0502; PMID:11677608
A:Accession: A10768
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-321 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD02471.1; PID:g16503337; GSPDB:GN00176
C:Genetics:
A:Gene: STY2320
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase
C:Keywords: isomerase

Query Match 57.4%; Score 944.5; DB 2; Length 321;
Best Local Similarity 58.5%; Pred. No. 7.4e-72;
Matches 182; Conservative 53; Mismatches 67; Indels 9; Gaps 1;

QY 7 KIFVAGHRLVGSIAVRKLOEQGFTNLVLTHTAELDLTROADVSEFFSQKPVY 66
DB 5 RIFVAGHRLVGSIAVRKLOEQGFTNLVLTHTAELDLTROADVSEFFSQKPVY 64
QY 67 KVGGIHANNTPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYKPAPOP 126
DB 65 KVGGIHANNTPADFIGVNLQIQTNVHSAVEHGVKLLFLGSSCIYKPAPOP 124

A:Gene: yefB
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 57.3%; Score 943.5; DB 1; Length 321;
Best Local Similarity 57.6%; Pred. No. 8.9e-72;
Matches 179; Conservative 56; Mismatches 67; Indels 9; Gaps 1;

QY 7 KIFVAGHRLVGSALVRKLEQGFNTLVKTHAELDLTRQADVESFFSQEKPVPYVILAAA 66
DB 5 RVFIAGHGMVGSALVRKLEQGFNTLVKTHAELDLTRQADVESFFSQEKPVPYVILAAA 64

QY 67 KVGGIHANNYPADFTGVNLIQTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 126
DB 65 KVGGIHANNYPADFTGVNLIQTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 124

QY 127 LTASLEPTNEWAIKAGIKTCQAYRIQHGWDATSGMPTNLYGPNDFPHNSHVLPAL 186
DB 125 LQGTLEPTNEPVAIAKIAGIKTCESYNRQGRDYRSVMPNTLYGPNDFPHNSHVLPAL 184

QY 187 MRRFHEAKVNGAEVVMVWGTSPLREFLHVDDLDADACVFLDDR-YS-----GLEHVN 237
DB 185 LRRFHDVENSPPNVVWVGSGTPKREFLHVDDLDADACVFLDDR-YS-----GLEHVN 237

QY 298 LSQTYDYLKLN 308
DB 305 LENTYNWFLEN 315

RESULT 8
A85829
fucose synthetase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Dec-2001
C:Accession: A85829
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85829
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-321 <STO>
A:Cross-references: GB:AB005174; NID:g12516221; PIDN:AAG57093.1; GSPDB:GN00145; UWGP:231
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 55.6%; Score 915.5; DB 2; Length 321;
Best Local Similarity 57.2%; Pred. No. 2e-69;
Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;

QY 7 KIFVAGHRLVGSALVRKLEQGFNTLVKTHAELDLTRQADVESFFSQEKPVPYVILAAA 66
DB 5 RVFIAGHGMVGSALVRKLEQGFNTLVKTHAELDLTRQADVESFFSQEKPVPYVILAAA 64

QY 67 KVGGIHANNYPADFTGVNLIQTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 126
DB 65 KVGGIHANNYPADFTGVNLIQTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 124

QY 127 LTASLEPTNEWAIKAGIKTCQAYRIQHGWDATSGMPTNLYGPNDFPHNSHVLPAL 186
DB 125 LQGTLEPTNEPVAIAKIAGIKTCESYNRQGRDYRSVMPNTLYGPNDFPHNSHVLPAL 184

QY 187 MRRFHEAKVNGAEVVMVWGTSPLREFLHVDDLDADACVFLDDR-YS-----GLEHVN 237
DB 185 LRRFHDVENSPPNVVWVGSGTPKREFLHVDDLDADACVFLDDR-YS-----GLEHVN 237

QY 298 LSQTYDYLKLN 308
DB 305 LENTYNWFLEN 315

RESULT 10
A98142
hypothetical 34.7K protein y4af [imported] - Agrobacterium tumefaciens (strain C58, Cer
Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: A98142
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: A98142
A:Status: preliminary
A:Molecule type: DNA

QY 238 IGSQGEVTTIRELAELVKEVGFEGKLGWDCTKPDGTPRKLMDSSKLASLGWTPKVSRLRG 297
DB 245 IGTGIDCTICELAETIAKVGVKGHTFTDTTKPDGAPRKLDDVTLHLQLGWNHKTILHKG 304

QY 298 LSQTYDYLKLN 308
DB 305 LENTYNWFLEN 315

RESULT 9
F90983
fucose synthetase [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
C:Accession: F90983
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90983
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-321 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA836261.1; PID:g1362306; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC82838
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 55.6%; Score 915.5; DB 2; Length 321;
Best Local Similarity 57.2%; Pred. No. 2e-69;
Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;

QY 7 KIFVAGHRLVGSALVRKLEQGFNTLVKTHAELDLTRQADVESFFSQEKPVPYVILAAA 66
DB 5 RVFIAGHGMVGSALVRKLEQGFNTLVKTHAELDLTRQADVESFFSQEKPVPYVILAAA 64

QY 67 KVGGIHANNYPADFTGVNLIQTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 126
DB 65 KVGGIHANNYPADFTGVNLIQTNVIHSAVEHGVKLLFLGSSCIYPKFAPQIPESAL 124

QY 127 LTASLEPTNEWAIKAGIKTCQAYRIQHGWDATSGMPTNLYGPNDFPHNSHVLPAL 186
DB 125 LQGTLEPTNEPVAIAKIAGIKTCESYNRQGRDYRSVMPNTLYGPNDFPHNSHVLPAL 184

QY 187 MRRFHEAKVNGAEVVMVWGTSPLREFLHVDDLDADACVFLDDR-YS-----GLEHVN 237
DB 185 LRRFHDVENSPPNVVWVGSGTPKREFLHVDDLDADACVFLDDR-YS-----GLEHVN 237

QY 238 IGSQGEVTTIRELAELVKEVGFEGKLGWDCTKPDGTPRKLMDSSKLASLGWTPKVSRLRG 297
DB 245 IGTGIDCTICELAETIAKVGVKGHTFTDTTKPDGAPRKLDDVTLHLQLGWNHKTILHKG 304

QY 298 LSQTYDYLKLN 308
DB 305 LENTYNWFLEN 315

RESULT 10
A98142
hypothetical 34.7K protein y4af [imported] - Agrobacterium tumefaciens (strain C58, Cer
Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: A98142
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: A98142
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-331 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK88659.1; PID:g15158384; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L185
A:Map position: linear chromosome
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 54.4%; Score 895.5; DB 2; Length 331;
Best Local Similarity 55.5%; Pred. No. 1e-67;
Matches 166; Conservative 59; Mismatches 73; Indels 1; Gaps 1;

QY 7 KIFVAGHGLVGSIAVRKLEQGFNLVLTAEHLDLTDQADVESFFSQEKPVYVILAA 66
DB 18 RVWVAGHTGWSALVRRLEREN-CEILKVSRLDLTRQYETEQQWMAARQVIVFAA 76
QY 67 KVGGIHANNYPADFGVNLQITQVNIHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 126
DB 77 KVGIIAANAAYPADFLYTNTLISNMIMKSAADIGVEKLLWMSGSCYIPKFAAQITENAL 136
QY 127 LTASLEPTNEWAIAKIAGIKTCQAYRIQHGMDAISGMPTNLGPNDFPHNSHVLPA 186
DB 137 LTGPLEPTNEWAIAKIALKLSQFYSIQVGLNCVSMPTNIYGLNDFDPOSSHVIPAM 196
QY 187 MRPFHAKVNGAEVVGWGTGSPLEFLHVDLADACVFLDRYSGLEHVNIGSGQEV 246
DB 197 IRRMHEAKISGQNKIVLWGTGSPLEFLHVDLADACVFLDRYSGLEHVNIGSGREIS 256
QY 247 RELAEIVKVEVGFEGKLGWDCPKDGTGPRKLMDSKSLASLGWTPKVSRLDGLSQT 305
DB 257 RNLAHLIAGIVGEGQIVFTSKPDGAPRKLDCSRLNALGWNSTVELRYGIQDLYEW 315

RESULT 11
AB3146
GDP-fucose synthetase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AB3146
R:Wood, D.W.; Stetubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AB3146
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAK45584.1; PID:g17743301; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: fcl
A:Map position: linear chromosome
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 54.4%; Score 895.5; DB 2; Length 331;
Best Local Similarity 55.5%; Pred. No. 1e-67;
Matches 166; Conservative 59; Mismatches 73; Indels 1; Gaps 1;

QY 7 KIFVAGHGLVGSIAVRKLEQGFNLVLTAEHLDLTDQADVESFFSQEKPVYVILAA 66
DB 18 RVWVAGHTGWSALVRRLEREN-CEILKVSRLDLTRQYETEQQWMAARQVIVFAA 76
QY 67 KVGGIHANNYPADFGVNLQITQVNIHSAHYEGVKLLFLGSSCIYPKFAPQIPESAL 126
DB 77 KVGIIAANAAYPADFLYTNTLISNMIMKSAADIGVEKLLWMSGSCYIPKFAAQITENAL 136
QY 127 LTASLEPTNEWAIAKIAGIKTCQAYRIQHGMDAISGMPTNLGPNDFPHNSHVLPA 186
DB 137 LTGPLEPTNEWAIAKIALKLSQFYSIQVGLNCVSMPTNIYGLNDFDPOSSHVIPAM 196

QY 187 MRPFHAKVNGAEVVGWGTGSPLEFLHVDLADACVFLDRYSGLEHVNIGSGQEV 246
DB 197 IRRMHEAKISGQNKIVLWGTGSPLEFLHVDLADACVFLDRYSGLEHVNIGSGREIS 256
QY 247 RELAEIVKVEVGFEGKLGWDCPKDGTGPRKLMDSKSLASLGWTPKVSRLDGLSQT 305
DB 257 RNLAHLIAGIVGEGQIVFTSKPDGAPRKLDCSRLNALGWNSTVELRYGIQDLYEW 315

RESULT 12
TI7792

hypothetical protein A295L - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: TI7792
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: TI7792
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-317 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAK96663.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: A295L
C:Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 53.6%; Score 882; DB 2; Length 317;
Best Local Similarity 54.6%; Pred. No. 1.3e-66;
Matches 167; Conservative 57; Mismatches 82; Indels 0; Gaps 0;

QY 3 DKSAKIFVAGHGLVGSIAVRKLEQGFNLVLTAEHLDLTDQADVESFFSQEKPVYV 62
DB 4 DKHSKIYVAGHTGWSLSVRLQKQYNTIITRSKELDLNQEQVTAFFEMETPEYVF 63
QY 63 LAAKVGGIHANNYPADFGVNLQITQVNIHSAHYEGVKLLFLGSSCIYPKFAPQIP 122
DB 64 LAAKVGGIHANNYPADFGVNLQITQVNIHSAHYEGVKLLFLGSSCIYPKESPNIK 123
QY 123 ESALLTASLEPTNEWAIAKIAGIKTCQAYRIQHGMDAISGMPTNLGPNDFPHNSHV 182
DB 124 EYLMTGFLPEPTNKPKYAIERIAEIMCDAYRKQFCNFSVMPNTLUSGNDRYDLQNGHV 183
QY 183 LPALMRPFHAKVNGAEVVGWGTGSPLEFLHVDLADACVFLDRYSGLEHVNIGSGQ 242
DB 184 FPLVLRKFYEAMINKVPSVKLWGTGIAREFLHVDDLARGLLIVMEKYNPEGPINIGYS 243
QY 243 EYVIRELAELVKEVGFEGKLGWDCPKDGTGPRKLMDSKSLASLGWTPKVSRLDGLSQT 302
DB 244 DVSISLAELVRDVGKAIYDTPMDGTCLKLIDSTKIKSLGWEPKISLIDNIKLV 303
QY 303 DMVLKN 308
DB 304 EDFTKN 309

RESULT 13
F97168

nucleoside-diphosphate-sugar epimerase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 31-Dec-2001
C:Accession: F97168
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80137.1; PID:g15025174; GSPDB:GN00168

A; Experimental source: Clostridium acetobutylicum ATCC824

C; Genetics:

A; Gene: CAC2179

C; Superfamily: Escherichia coli probable UDPglucose 4-epimerase

Query Match 53.3%; Score 877.5; DB 2; Length 309;

Best Local Similarity 53.6%; Pred. No. 3.1e-66;

Matches 164; Conservative 66; Mismatches 75; Indels 1; Gaps 1;

```
QY 4 KSAKIFVAGHRLVGSNAIVRKLQEQFTNLVLTAEHLDTQADVESFQSKPVYVL 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 EDSKIYIAGHTGVGSAILRLNRRGYKXNVVRTHKELDLMHQESVKKFLERKPYVVL 62

QY 64 AAQVGGIHHANTYPADFIGVNLQIOTNVTHSAVEHGVKKLLFLGSSCIYKPAPOIPE 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 SAAKVGGIQNNISNPVDFLMDNLIIENYVTKSFVGIENLLFLGSSCIYKPAPOPLKE 122

QY 124 SALLTASLEPTNEWYAIAGIKTQOAYRIQHGWDIAISGMPTNLYGPNDFHPENSHVLP 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 EYLLSGYLEPTNEGAIKISGLKMCYYSKQGLNYISAMPENLYGMRDNFDLKTSHVM 182

QY 184 PALMRPFHAKVNGAEVVMVGTGSPLRFLHVDDLDACVFLLDYRSGLE-HVNIGSQ 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 AALIRPFHAKVSGSQEISWGSQEYREFTYIEDLADGIIIFLMEHGEKVGFNLITGCGK 242

QY 243 EVTIRELAELVEWGVFEGKLGWDCTKPDGTPRKLMDSSKLASLGWTPKVSRLDGLSQT 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 DIKIKDLAYKIDVGVFGKGNIFDKSKPDGMFRKMDVSKNSLGHWHYKVELDEGITKY 302

QY 303 DWYLNK 308
   |||
Db 303 RMYLNN 308
```

RESULT 14

DB1288 probable fucose synthetase Cj1428c [imported] - Campylobacter jejuni (strain NCTC 11168)

C; Species: Campylobacter jejuni

C; Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C; Accession: DB1288

R; Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr

Nature 403, 665-668, 2000

A; Title: The genome sequence of the food-borne pathogen Campylobacter jejuni. reveals hyp

A; Reference number: AB1250; MUID:20150912; PMID:10688204

A; Accession: DB1288

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-346 <PAR>

A; Cross-references: GB:AL111168; NID:G6968723; PIDN:CAB73852.1; PID:G696885

A; Experimental source: serotype O2, strain NCTC 11168

C; Genetics:

A; Gene: fcl; Cj1428c

Query Match 41.7%; Score 686; DB 2; Length 346;

Best Local Similarity 42.6%; Pred. No. 4.8e-50;

Matches 145; Conservative 53; Mismatches 102; Indels 40; Gaps 4;

```
QY 5 SAKIFVAGHRLVGSNAIVRKLQEQFTNLVLTAEHLDTQADVESFQSKPVYVILA 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 NSKIYIAGHKGTAGTALVENLQKGFNNLVLTQRLDLVNOQAAVAKFFKEEKPEYVFLT 63

QY 65 AAQVGGIHHANTYPADFIGVNLQIOTNVTHSAVEHGVKKLLFLGSSCIYKPAPOIPES 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 AVLPGA-ANVAQRADFIYENLMQNNVTHNSFLNNVKLVFGSGYMPENAKNPLKEE 122

QY 125 ALLTASLEPTNEWYAIAGIKTQOAYRIQHGWDIAISGMPTNLYGPNDFHPENSHVLP 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 YLFQGDLEYGAYSGAAGIAGIMCESYNIQVGTNFIILVNLNLYGTKANFDFGKSRVLP 182

QY 185 ALMRPFH- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 ALLRKPHLAKLISEGNITQIQLDKNNPFEAKKEYLHNFGLSKSVETMGTKVRREFTH 242
```

```
QY 216 VDDLADACVFL- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 SDDLADVAIYTMQNDPDKDLIKDRKSKNTHINIGIDYSIKEVALMVKNIYVFGSGELVF 302

QY 266 DCTKPDGTPRKLMDSSKLASLGWTPKVSRLDGLSQTQYDWY 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 NTSRDPSTMDRLMDCSKIHSLGKHKIBLKDKGIKMWEY 342
```

RESULT 15

AD3615

gdp-4-dehydro-D-rhamnose reductase (EC 1.1.1.187) [imported] - Brucella melitensis (str)

C; Species: Brucella melitensis

C; Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C; Accession: AH3615

R; DelVecchio, V.G.; Kaparal, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,

; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes,

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A; Title: The genome sequence of the facultative intracellular pathogen Brucella meliten

A; Reference number: AD3252; PMID:11756688

A; Accession: AH3615

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-246 <KUR>

A; Cross-references: GB:AE008918; PIDN:AAL54091.1; PID:gi7985049; GSPDB:GN00191

A; Experimental source: strain 16M

C; Genetics:

A; Gene: BME110849

A; Map position: II

C; Keywords: Oxidoreductase

Query Match 36.2%; Score 596.5; DB 2; Length 246;
Best Local Similarity 53.2%; Pred. No. 1e-42; Mismatches 38; Indels 21; Gaps 4;

```
QY 7 KIFVAGHRLVGSNAIVRKLQEQFTNLVLTAEHLDTQADVESFQSKPVYVILAAA 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 19 KIFVAGHTGVGSAILRLQHEH-CDIITAHSVLDTLQGTPTNFISGHRPDVILIAA 77

QY 67 KVGGIHANTYPADFIGVNLQIOTNVTHSAVEHGVKKLLFLGSSCIYKPAPOIPESAL 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 RVGGILANSRFPADFLYNNLAIGNLIHAHQIGVERLLWLGLSSCIYPRDAAQPLTEDAL 137

QY 127 LTASLEPTNEWYAIAGIKTQOAYRIQHGWDIAISGMPTNLYGPNDFHPENSHVLPAL 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 LTGPLEPTNEAYAIAGIKTQOAYRIQHGWDIAISGMPTNLYGPNDFHPENSHVLPAL 197

QY 187 MRRPFHAKVNGAEV- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 IRRVHEARMRGAEEVCFGAAASPCGNSCMW-TTSPMRACI- - - - - - - - - - - - - - - - - -
```

Search completed: June 3, 2003, 05:09:20

Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 05:02:36 ; Search time 21 Seconds
(without alignments)
437.141 Million cell updates/sec

Title: US-10-089-014-1
Perfect score: 1645
Sequence: 1 MDKSAKIFVAGHGLVGS.....SLRDGLSQTYDWLVKNVCR 312

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	463.5	28.2	321	2	US-08-937-972-3
2	402	24.4	271	2	US-08-937-972-6
3	206.5	12.5	3782	4	US-09-105-537-4
4	203.5	12.4	329	4	US-09-036-987A-26
5	203.5	12.4	329	4	US-09-370-700-26
6	203.5	12.4	337	3	US-09-320-878-15
7	199.5	12.1	337	4	US-09-105-537-14
8	190.5	11.6	325	4	US-09-194-905-11
9	180	10.9	278	4	US-09-149-476-701
10	169.5	10.3	348	4	US-09-113-536-2
11	169.5	10.3	348	4	US-09-183-183-2
12	169.5	10.3	348	5	PCT-US95-05785-2
13	151.5	9.2	306	4	US-09-392-772-6
14	151.5	9.2	344	4	US-09-392-772-2
15	139	8.4	427	4	US-09-392-772-4
16	129.5	7.9	322	2	US-08-576-626A-33
17	129.5	7.9	2544	2	US-08-576-626A-32
18	129	7.8	327	3	US-09-154-874-8
19	125	7.6	366	4	US-09-638-715-2
20	125	7.6	366	4	US-09-638-715-4
21	124	7.5	337	2	US-08-923-856-4
22	124	7.5	337	3	US-09-216-294-4
23	122.5	7.4	326	3	US-09-154-874-9
24	107	6.5	153	4	US-09-392-772-8
25	106	6.4	299	2	US-08-576-626A-35
26	104	6.3	338	3	US-08-722-184-8
27	104	6.3	338	4	US-09-043-937A-12

28	100	6.1	335	3	US-08-722-184-2	Sequence 2, Appli
29	100	6.1	335	4	US-09-043-937A-6	Sequence 6, Appli
30	96.5	5.9	290	2	US-08-576-626A-34	Sequence 34, Appl
31	95	5.8	335	3	US-08-722-184-10	Sequence 10, Appl
32	95	5.8	335	4	US-09-043-937A-14	Sequence 14, Appl
33	95	5.8	1722	4	US-09-194-612A-1	Sequence 1, Appli
34	94.5	5.7	342	3	US-08-722-184-6	Sequence 6, Appli
35	94.5	5.7	342	4	US-09-043-937A-10	Sequence 10, Appl
36	90.5	5.5	338	3	US-08-722-184-4	Sequence 4, Appli
37	90.5	5.5	338	4	US-09-043-937A-8	Sequence 8, Appli
38	90	5.5	320	4	US-09-134-001C-4439	Sequence 4439, Ap
39	89.5	5.4	371	4	US-09-043-937A-4	Sequence 4, Appli
40	89	5.4	181	4	US-09-194-905-6	Sequence 6, Appli
41	89	5.4	924	4	US-09-267-311-2	Sequence 2, Appli
42	88.5	5.4	1447	4	US-09-041-886-25	Sequence 25, Appl
43	88.5	5.4	1447	5	PCT-US94-05277-2	Sequence 2, Appli
44	86	5.2	312	4	US-09-475-316A-108	Sequence 108, App
45	85	5.2	320	2	US-08-846-762-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-937-972-3
; Sequence 3, Application US/08937972
; Patent No. 5932443
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: ANTIGENS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,972
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0400 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 321 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BLADNOT04
; CLONE: 1318190
US-08-937-972-3

Query Match 28.2%; Score 463.5; DB 2; Length 321;

	Best Local Similarity	35.4%; Pred. No. 4.8e-42;
	Matches	111; Conservative
	Gaps	64; Mismatches
	Indels	15; Gaps
Qy	5	SAKIFVAGHRLGVSASIVRKLQEQQF---TNLVLKTHAELDLTRQADVESFFSQEKPVVY 60
Dd	7	SMRILVTGGSLGVKAII-QKVVDAGAGLPGEDWVFVSSKOADLTDTACTRALFEKVQPHT 65
Qy	61	VILAALKVGGIHANNTYPADFIGNLOIQTNVIHSAYEHGVKKLLFLGSSCIYPKPAPQP 120
Dd	66	VIHLAAVMVGGFLRNKIKNLDIPWRKNVHMNDMLNLSAFEVGAARKVCVSCLSCTCFDPKTTP 125
Qy	121	IPESALLITASLEPTNEWATAKIAGIKTCQAYRTQHGWDAISGMTPLNYLGNDNFHPENS 180
Dd	126	IDETMIHGPHNSFCYSYAKRMIDVNRAIFYOYGCTFAVIPFTNVFGPHDNFNLEDG 185
Qy	181	HVLPALMRRPHEAKNGAEVWWGTSPUREFLIHDVDDACAVFLDLYSGLEHV--NI 238
Dd	186	HVLPLGLHKVHLAKSSGS-ALTVMGTGNPRQFIYSLDLAQLFITWLIREYNEVEPIILSV 244
Qy	239	GSQGERTIRELAELAVEKVPGEKLGWDCTKPDGTPRKLMDSKLAS----LGMTPKVS L 294
Dd	245	GEDEVSIAKAAEAVAEMDFHGEVTFDTTKSDCQFKKTASNKLRITYLPDRFPETP---P 301
Qy	295	RDLQSOTDYWLXKN 308
Dd	302	KOARVKETCAWFNTN 315

RESULT 2
US-08-937-972-6

```

; Sequence 6, Application US/08937972
; Patent No. 5932443
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: ANTIGENS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,972
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0400 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 271 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

```


APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 30622002120
CURRENT APPLICATION NUMBER: US/09/320,878A
CURRENT FILING DATE: 1999-05-27 OF 09/141,908
EARLIER APPLICATION NUMBER: CIP OF 09/141,908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073,538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846,247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 15
LENGTH: 337
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-320-878-15

Query Match 12.4%; Score 203.5; DB 3; Length 337;
Best Local Similarity 24.5%; Pred. No. 1e-13;
Matches 84; Conservative 49; Mismatches 139; Indels 71; Gaps 11;
QY 7 KIFVAGHRLVGSIAVRKLEQGFNTLVKTHAELD-LT-----44
DB 2 RLLVTGGAGFIGSHFVRQLLAGAYPDVPADEVILDSLTAYAGNRANLAPVDADPRLRFVH 61
QY 45 -----ROADVESFSEKPYVIVLAAGKVGIIHANNTPYADFIGVNLQIQTN 91
DB 62 GDIRDAGLLARELGRGDAIVHFAAESHVDRSIAGASV-----FTETNVQGTOT 109
QY 92 VIHSAYEHGVKKLLFLGSSCIYKPAQPIPESSALLTASLEPTNEWYAIKAGIKTCA 151
DB 110 LLOCAVDAGVRVHVSTDEVYSGDSWTES-----SPLP-NSPYAASKAGSLVARA 164
QY 152 YRIQHWDAISGMPTNLYGNPNFHPENSHVLPALMRPFHEAKVNGAEVVMVGTGSPUR 211
DB 165 YHRTYGLDVRITRCNNYGPYQ--HPEK-----LIPLFVTNLLDGG-TLPLYGDCANVR 215
QY 212 EFLHVDLADACVFLLDRLYSGLEHVNIGSQEVTIRELAELVKEVVGFEKLGWCTKPD 271
DB 216 EWHTDDHCRGIALVLAGGRAGEIYHIGGLELTNRELTGILLDSLGAD-----WSSVRKV 271
QY 272 GTP-----RKLMDSSKL-ASLGWTPKVSRLDGLSOTYDWYLNK 308
DB 272 ADKRGHDLRYSLDGGKIERELGYRQVSPADGLARTVRYREN 314

RESULT 7
US-09-105-537-14
Sequence 14, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 337
TYPE: PRT
ORGANISM: Streptomyces venezuelae

US-09-105-537-14
Query Match 12.1%; Score 199.5; DB 4; Length 337;
Best Local Similarity 24.2%; Pred. No. 2.8e-13;
Matches 83; Conservative 50; Mismatches 139; Indels 71; Gaps 11;
QY 7 KIFVAGHRLVGSIAVRKLEQGFNTLVKTHAELD-LT-----44
DB 2 RLLVTGGAGFIGSHFVRQLLAGAYPDVPADEVILDSLTAYAGNRANLAPVDADPRLRFVH 61
QY 45 -----ROADVESFSEKPYVIVLAAGKVGIIHANNTPYADFIGVNLQIQTN 91
DB 62 GDIRDAGLLARELGRGDAIVHFAAESHVDRSIAGASV-----FTETNVQGTOT 109
QY 92 VIHSAYEHGVKKLLFLGSSCIYKPAQPIPESSALLTASLEPTNEWYAIKAGIKTCA 151
DB 110 LLOCAVDAGVRVHVSTDEVYSGDSWTES-----SPLP-NSPYAASKAGSLVARA 164
QY 152 YRIQHWDAISGMPTNLYGNPNFHPENSHVLPALMRPFHEAKVNGAEVVMVGTGSPUR 211
DB 165 YHRTYGLDVRITRCNNYGPYQ--HPEK-----LIPLFVTNLLDGG-TLPLYGDCANVR 215
QY 212 EFLHVDLADACVFLLDRLYSGLEHVNIGSQEVTIRELAELVKEVVGFEKLGWCTKPD 271
DB 216 EWHTDDHCRGIALVLAGGRAGEIYHIGGLELTNRELTGILLDSLGAD-----WSSVRKV 271
QY 272 GTP-----RKLMDSSKL-ASLGWTPKVSRLDGLSOTYDWYLNK 308
DB 272 ADKRGHDLRYSLDGGKIERELGYRQVSPADGLARTVRYREN 314

RESULT 8
US-09-194-905-11
Sequence 11, Application US/09194905
Patent No. 6306627
GENERAL INFORMATION:
APPLICANT: DECKER, Heinrich
TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
TITLE OF INVENTION: GLA.O AND THEIR USE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194,905
FILING DATE: 29-JUL-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/EP97/02826
FILING DATE: 30-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 19622783.6
FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Granados, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 026083/0193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 325 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-194-905-11

Query Match 11.6%; Score 190.5; DB 4; Length 325;
Best Local Similarity 25.4%; Pred. No. 2.6e-12;
Matches 85; Conservative 50; Mismatches 136; Indels 63; Gaps 14;

QY 8 IFVAGHRLGVGAIVKRLQEQG-----FTNLVLKTHA-----ELDL 43
DB 5 ILVTGGAGIRAYVRLLSPAGGVAVTLDKLTYAGSLARLHRAVRDHPGLTFVQGDV 64
QY 44 TRQADVSPFSQKPVVYILAAKVGIIHANNYPADFTGVNLQIQTNVIHSAYEHGVKK 103
DB 65 CDTALVDTLAARHDDIVFAESHV---DRSITDSGAFRTNVLTGTVQLDAAALRHGVRT 121
QY 104 LFLGSSCIYKFAQPIPESALLTAS-LEPTNEWYAIKAGIKTCQAYRIQHGWDALS 162
DB 122 FHVSTDEVYGS-----LPHGAASDPLLP7SP-YAASKAASDLMALAHRTHTGLDVRV 175
QY 163 GMPNTLYGNDFNPENSHVLPALMRPFHEAKVNGAEVVVWGTGSPLEFLHVDLADA 222
DB 176 TRCSNNFGPHQ--HPEK-----LIPRFLTSLSGG-TVPLYDGDGHVHDWLHVDHVR 226
QY 223 CVFLDLRYSL--EHNIGSGQEVITRELAELV-----REVVGFEKGLGMDCTKPDG 272
DB 227 VE--LVRVSGRPGELIYNGGTSPLNLELTHRLALCGAGPERIVHVENRKGH----- 278
QY 273 TPKLMDSKL-ASLWTKVSLRDLGSLQTYDWY 305
DB 279 -RRYAVDHSKITAEGLYRRTDFATALADTAKWY 311

RESULT 9
US-09-149-476-701
Sequence 701, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 04:00:21 ; Search time 54 Seconds
(without alignments)
769.892 Million cell updates/sec

Title: US-10-089-014-1

Perfect score: 1646

Sequence: 1 MSDKSAKIFVAGHGLVGS.....SLRDGLSQTYDWYLNKNCVR 312

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1646	100.0	312	22	AAG62614
2	1515	92.0	314	21	AAV54114
3	1353	82.2	299	21	AAG09919
4	1353	82.2	307	21	AAG09918
5	1353	82.2	359	21	AAG09917
6	1241	75.4	310	22	AAW00095
7	971	59.0	339	18	AAW21770
8	971	59.0	339	18	AAW21771
9	943.5	57.3	317	21	AAV81360
10	937.5	57.0	321	20	AAV29659

11	937.5	57.0	321	21	AAV54115	A GDP-4-keto-6-deo
12	915.5	55.6	323	20	AAW88319	E. coli O157 anti
13	718.5	43.7	576	22	ABG18368	Novel human diag
14	671	40.8	162	22	AAW00029	GDP-4-keto-6-deoxy
15	593.5	36.1	307	20	AAW88305	E. coli O111 anti
16	550.5	33.4	597	22	ABG28981	Novel human diag
17	550.5	33.4	693	22	ABG25460	Novel human diag
18	500	30.4	244	22	ABG18364	Novel human diag
19	468.5	28.5	321	23	ABG34139	Antibody productio
20	463.5	28.2	321	20	AAV28286	Amino acid sequenc
21	463.5	28.2	321	21	AAV54116	A GDP-4-keto-6-deo
22	451.5	27.4	314	21	AAV81361	Human GDP-fucose s
23	437.5	26.6	321	22	ABV59947	Drosophila melanog
24	430	26.1	260	22	ABG18365	Novel human diag
25	425	25.8	1165	22	ABG00133	Novel human diag
26	381	23.1	157	22	ABG28975	Novel human diag
27	362.5	22.0	545	22	ABG13190	Novel human diag
28	337.5	20.5	214	18	AAW20804	H. pylori cytoplas
29	324	19.7	455	22	AAU28305	Novel human secret
30	313.5	19.0	325	21	AAW07566	Protein encoded by
31	271	16.5	380	21	AAG35197	zea mays protein f
32	271	16.5	420	21	AAG35196	zea mays protein f
33	263.5	16.0	377	21	AAW06774	Arabidopsis thalia
34	263.5	16.0	377	21	AAW14539	Arabidopsis thalia
35	263.5	16.0	423	21	AAW14538	Arabidopsis thalia
36	263.5	16.0	426	21	AAW06773	Arabidopsis thalia
37	262.5	15.9	336	21	AAW83788	S. fradiae tylosin
38	253	15.4	308	22	AAW96739	Putative P. abyss
39	249	15.1	311	22	AAW90108	C glutamicum prote
40	249	15.1	311	22	AAW79385	Corynebacterium gl
41	249	15.1	311	22	AAW79467	Corynebacterium gl
42	247	15.0	314	21	AAW35198	zea mays protein f
43	240	14.6	226	18	AAW33273	S. fradiae tylosin
44	239.5	14.6	313	21	AAW06775	Arabidopsis thalia
45	239.5	14.6	313	21	AAW14540	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAG62614
ID AAG62614 standard; protein; 312 AA.
XX AC
XX AAG62614;
XX 06-SEP-2001 (first entry)
XX DT
XX DE A thaliana GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase.
XX KW GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase; GDP-L-fucose;
XX KW sugar chain.
XX OS Arabidopsis thaliana.
XX PN WO200138507-A1.
XX PD 31-MAY-2001.
XX PF 30-MAR-2000; 2000WO-JP02049.
XX PR 19-NOV-1999; 99JP-0329045.
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
XX PI Nakayama K, Jigami Y;
XX DR WPI; 2001-381292/40.
XX DR N-PSDB; AA45734.
XX PT Recombinant GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase
PT derived from Arabidopsis for efficient production of GDP (guanine
PT diphosphate)-L-fucose in vitro or in vitro

XX PS Claim 1; Page 24-26; 35pp; Japanese.

XX CC The present invention provides the protein and coding sequences of the

XX CC Arabidopsis thaliana GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-

XX CC reductase. The sequence can be used for the efficient production of

XX CC GDP-L-fucose, which is essential to the production of functionally

XX CC important sugar chains containing L-fucose. The present sequence is the

XX CC protein of the invention.

XX SQ Sequence 312 AA;

Query Match 100.0%; Score 1646; DB 22; Length 312;

Best Local Similarity 100.0%; Pred. No. 1.5e-159;

Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLTAEHDLTRQADVSEFSSQKPY 60

DB 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLTAEHDLTRQADVSEFSSQKPY 60

QY 61 VILAAKVGGIHANNTYPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYKPAOP 120

DB 61 VILAAKVGGIHANNTYPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYKPAOP 120

QY 121 IPESALLTASLEPTNEWYAIAGIKTCQAYRIQHGWDATSGMPTNLGPNDFHPENS 180

DB 121 IPESALLTASLEPTNEWYAIAGIKTCQAYRIQHGWDATSGMPTNLGPNDFHPENS 180

QY 181 HVLPMRRFHEAKVNGAEVVGWGTSPRLREFLHVDDLDADACVFLLDRIQ 240

DB 181 HVLPMRRFHEAKVNGAEVVGWGTSPRLREFLHVDDLDADACVFLLDRIQ 240

QY 241 GQVVTIRELAELVKEVVGEGKLGWDTKPDGTPRKLMDSSKLASLGWTPKVSRLDLSQ 300

DB 241 GQVVTIRELAELVKEVVGEGKLGWDTKPDGTPRKLMDSSKLASLGWTPKVSRLDLSQ 300

QY 301 TYDWYLNKVCNR 312

DB 301 TYDWYLNKVCNR 312

RESULT 2

AA54114

ID AAY54114 standard; Protein; 314 AA.

AC AAY54114;

XX 27-MAR-2000 (first entry)

DE A GDP-4-keto-6-deoxy-D-mannose epimerase/reductase protein.

XX GDP-4-keto-6-deoxy-D-mannose epimerase/reductase; GDP-D-mannose;

XX GDP-L-galactose; vitamin C; ascorbic acid; L-ascorbic acid;

XX ascorbic acid pathway enzyme; hexokinase; glucose phosphate isomerase;

XX phosphomannose isomerase; phosphomannomutase; L-galactose dehydrogenase;

XX GDP-D-mannose pyrophosphorylase; GDP-D-mannose:GDP-L-galactose epimerase;

XX GDP-L-galactose phosphorylase; L-galactose-1-P-phosphatase;

XX L-galactono-gamma-lactone dehydrogenase; ester.

XX Arabidopsis thaliana.

XX WO9964618-A1.

XX 16-DEC-1999.

XX 26-MAY-1999; 99WO-US11576.

XX 08-JUN-1998; 98US-0088549.

XX 17-MAR-1999; 99US-0125073.

XX 18-MAR-1999; 99US-0125054.

XX (DCVB-) DCV INC DBA BIO-TECH RESOURCES.

XX

PI Berry A, Running JA, Severson DK, Burlingame RP;

XX WPI: 2000-105890/09.

DR N-PSDB; AA245315.

XX Production of ascorbic acid or esters, using microorganisms or plants

PT which have genetic modification in enzymes involved in the ascorbic

PT acid synthesis pathway -

XX Claim 27; Page 167-168; 187pp; English.

XX The present sequence represents a GDP-4-keto-6-deoxy-D-mannose epimerase/

CC reductase. The enzyme catalyses the conversion of GDP-D-mannose to

CC GDP-L-galactose. The enzyme can be modified, and used to produce

CC transgenic microorganisms, which can be used in fermentation techniques

CC to produce vitamin C (ascorbic acid, L-ascorbic acid). The enzyme is

CC modified to increase its action. Other ascorbic acid pathway enzymes

CC which may be used in the method of the invention include hexokinases,

CC glucose phosphate isomerases, phosphomannose isomerases,

CC phosphomannomutases, GDP-D-mannose pyrophosphorylases,

CC GDP-D-mannose:GDP-L-galactose epimerases, GDP-L-galactose phosphorylases,

CC L-galactose-1-P-phosphatases, L-galactose dehydrogenases, and

CC L-galactono-gamma-lactone dehydrogenases. The methods can be used for

CC the production of ascorbic acid or esters using microorganisms or plants.

XX SQ Sequence 314 AA;

Query Match 92.0%; Score 1515; DB 21; Length 314;

Best Local Similarity 92.5%; Pred. No. 3.7e-146;

Matches 298; Conservative 1; Mismatches 5; Indels 19; Gaps 5;

QY 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLTAEHDLTRQADVSEFSSQKPY 60

DB 1 MSDKSAKIFVAGHRLVGSIVRKLQEQGFTNLVLTAEHDLTRQADVSEFSSQKPY 60

QY 61 VILAAKVGGIHANNTYPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYKPAOP 120

DB 61 VILAAKVGGIHANNTYPADFIGVNLQIQTNVHSAYEHGVKLLFLGSSCIYKPAOP 120

QY 121 IPESALLTASLEPTNEWYAIAGIKTCQAYRIQHGWDATSGMPTNLGPNDFHPENS 180

DB 121 IPESALLTASLEPTNEWYAIAGIKTCQAYRIQHGWDATSGMPTNLGPNDFHPENS 179

QY 181 HVLPMRRFHEAKVNGAEVVGWGTSGS-----PLR-BFLHVDDLDADACVFLLDRIQ 231

DB 180 HVLPMRRFHEAKV-----WSGSGGVGVKVPFEGKFLHVDDLDADACVFLLDRIQ 232

QY 232 -GLEHVNIGSGQVVTIRELAELVKEVVGEGKLGWDTKPDGTPRKLMDSSKLASLGWTP 290

DB 233 RGLEHVNIGSGQVVTIRELAELVKEVVGEGKLGWDTKPDGTPRKLMDSSKLASLGWTP 292

QY 291 KVSRLDGLSQTVDWYLNKVCNR 312

DB 293 KVSRLDGLSQTVDWYLNKVCNR 314

RESULT 3

AA09919

ID AAG09919 standard; Protein; 299 AA.

XX AAG09919;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 8036.

XX Protein identification; signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX	06-SEP-2000.		PR	02-JUL-1999;	99US-0142055.
PD			PR	06-JUL-1999;	99US-0142390.
XX			PR	08-JUL-1999;	99US-0142803.
PF	25-FEB-2000; 2000EP-0301439.		PR	09-JUL-1999;	99US-0142920.
XX			PR	12-JUL-1999;	99US-0142977.
PR	25-FEB-1999;	99US-0121825.	PR	13-JUL-1999;	99US-0143542.
PR	05-MAR-1999;	99US-0123380.	PR	14-JUL-1999;	99US-0143624.
PR	09-MAR-1999;	99US-0123548.	PR	15-JUL-1999;	99US-0144005.
PR	23-MAR-1999;	99US-0125788.	PR	16-JUL-1999;	99US-0144085.
PR	25-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135829.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.

```

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.2%; Score 1353; DB 21; Length 299;
Best Local Similarity 89.5%; Pred. No. 1.2e-129;
Matches 255; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSDKSAKIFVAGHRLVGSALVRKLOEQFTNLVLKTHAELDLTRQADVESFFSOEKPY 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 7 MLEKSAKIFVAGHRLVGSALVRKLOEQFTNLVLKTHAELDLTRQADVESFFATEKPY 66
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 VILAAKVGGIHANNTYPADFIGVNLQIQTNVIHSAIEHGVKKLLFLGSSCIYKPAQP 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 67 VILAAKVGGIHANNTYPADFIGVNLQIQTNVIHSAIEHGVKKLLFLGSSCIYKPAQP 126
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 IPESALLTASLEPTNEWYAIKAGIKTCQAVRIQHWDAISGMPTNLGPNDFHPENS 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 127 IPESALLTGLEPTNEWYAIKAGIKTCQAVRLQHWDAISGMPTNLGQDNDFHPENS 186
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 HVLPALMRPFHEAKVNGASEVWVGSGPLREFLHVDDLDACVFLLDYRSGLEHVNIGS 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 HVLPALMRPFHEAKANNANDEVWVGSGPLREFLHVDDLDACVFLMDQYSGFEHVNGS 246
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 241 GQEVITRELAELVKEVVGEGKLGWDCITKPDGTPRKLMDSSKLAS 285
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 247 GVEVTIKELAEVLVKEVVGKGLVMDTTKPDGTPRKLMDNSKLAS 291
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 4
AAG09918
ID AAG09918 standard; Protein; 307 AA.
AC
XX AAG09918;
XX
DT 17-OCT-2000 (first entry)
XX

```

```

DE Arabidopsis thaliana protein fragment SEQ ID NO: 8035.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.

```

PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142377.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 1353; DB 21; Length 307;

Best Local Similarity 89.5%; Pred. No. 1.3e-129;

Matches 255; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSDSAKIFVAGHRLVGSIAIVRKLOEQFTNLVLTTHAELDLTRQADVESFSEKPYV 60
DB 15 MLESAKIFVAGHRLVGSIAIVRKLOEQFTNLVLTTHAELDLTRQADVESFSEKPYV 74

QY 61 VILAAKGGIHNNTYPADFIGVNLQIOTNVIHSAYEHGVKLLFLGSSCIYPKFAPQ 120
DB 75 VILAAKGGIHNNTYPADFIGVNLQIOTNVIHSAYEHGVKLLFLGSSCIYPKFAPQ 134

QY 121 IPESALLTASLEPTNEWAIAKIAGIKTCQAYRIQHGWDIAISGMPTNLYGPNDFHPENS 180
DB 135 IPESALLTGPLEPTNEWAIAKIAGIKTCQAYRIQHGWDIAISGMPTNLYGPNDFHPENS 194

QY 181 HVLPALMRPFHEAKUNGABEVVWGTGSPRLRFHLVDDLDADACVFLLDYSLGHEVNVGS 240
DB 195 HVLPALMRPFHEAKUNADDEVVWGTGSPRLRFHLVDDLDADACVFLMDQYSGFHVNVGS 254

QY 241 GOEVTIRELAELVKVGVGEGKLGWDCTKPDGTPRKLMDSKIAS 285

DB 255 GVEVTIKELAEVLKVGKLVGWDTKPDGTPRKLMDSKIAS 299

RESULT 5
AAG09917
ID AAG09917 standard; Protein; 359 AA.
XX AC AAG09917;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 8034.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154779.
 PR 20-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162142.

Query Match 82.2%; Score 1353; DB 21; Length 359;
 Best Local Similarity 89.5%; Pred. No. 1.6e-129;
 Matches 255; Conservative 16; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSDKSAKIFVAGHGLVGSIVRKLQEQFTNLVLKTHAELDLTROADVESFFSOEKPYV 60
 Db 67 MLEKSAKIFVAGHGLVGSIVRKLQEQFTNLVLKTHAELDLTROADVESFFSOEKPYV 126

QY 61 VILAAKVGGIHANNYPADFTGVNLQIOTNVIHSAHYGVKKLLPLGSSCIYKFPAPOP 120.
 Db 127 VILAAKVGGIHANNYPADFTGVNLQIOTNVIHSAHYGVKKLLPLGSSCIYKFPAPOP 186

QY 121 IPESALLTASLPTNEWYAIKAGIKTCQAVRIOHGWDAISGMPTNLGPNNDHPENS 180
 Db 187 IPESALLTGLPFTNEWYAIKAGIKTCQAVRIOHGWDAISGMPTNLGPNNDHPENS 246

QY 181 HVLPALMRERHEAKVNGAEVWVGTSPLRFLHVDDIADACVFLLDYRSGLEHVNIGS 240
 Db 247 HVLPALMRERHEAKVNGAEVWVGTSPLRFLHVDDIADACVFLMDQYSGFEHVNIGS 306

QY 241 GOEVTIRLAEALVKEVVGEGKLGWDCPKDCTPRKLMDSKSLAS 285
 Db 307 GVEVTIKELAEALVKEVVGEGKLGWDCPKDCTPRKLMDSKSLAS 351

RESULT 6
 AAM00095
 ID AAM00095 standard; Protein; 310 AA.
 AC AAM00095;
 XX
 DT 28-SEP-2001 (first entry)
 XX
 DE GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase-4-reductase sequence #163.
 XX
 KW Moss; carbohydrate metabolism related protein; CMRP; sugar; cofactor;
 fine chemical production; carbohydrate; polysaccharide.
 XX
 OS Physcomitrella patens.
 XX
 PN WO200144476-A2.
 XX
 PD 21-JUN-2001.
 XX
 PF 14-DEC-2000; 2000WO-EP12697.
 XX
 PR 16-DEC-1999; 99US-0171101.
 XX
 PA (BADI) BASF PLANT SCI GMBH.
 XX
 PI Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;
 PI Frank M, Freund A, Duwenig E, Schmidt R, Reski R;
 XX
 DR WPI; 2001-398155/42.
 N-PSDB; AAH88781.
 XX
 PT Novel moss nucleic acid molecules encoding a carbohydrate metabolism
 related protein useful for modulating production of fine chemicals such
 as carbohydrates, cofactors and enzymes from microorganisms and plants

Claim 30; Page 127; 133pp; English.
 This invention relates to nucleic acid molecules AAH88708 - AAH88796 isolated from Physcomitrella patens (a moss), which encode carbohydrate metabolism related proteins (CMRP) represented in AAM00022 - AAM00110. Included in the invention is a vector containing the CMRP cDNA, and a host cell transformed with the vector. The host cell (a microorganism, Corynebacterium or Brevibacterium, moss or algae or a plant cell) is useful for producing a fine chemical such as carbohydrates, cofactors and/or enzymes. The nucleic acid molecules are suitable for modifying a carbohydrate production system in a host, e.g., microorganisms and plants. They are also useful to identify those DNA sequences and enzymes in other species which are useful to modify the biosynthesis of starch, cell wall polysaccharides and sugars. The nucleic acid molecules may be utilised in the genetic engineering of Corynebacterium glutamicum and the related Brevibacterium species and Acetobacter xylinum and Chlorella to make it a better or more efficient producer of one or more fine chemicals. Mutagenesis of one or more CMRPs may also result in CMRPs having altered activities which indirectly impact the production of one or more desired fine chemicals from plants. Primers AAH88705 - AAH8707 are used in the sequencing of the CMRP cDNA sequences of the invention.

Sequence 310 AA;

Query Match 75.4%; Score 1241; DB 22; Length 310;
 Best Local Similarity 75.5%; Pred. No. 3.7e-118;
 Matches 231; Conservative 37; Mismatches 36; Indels 2; Gaps 1;

Claim 1; Page 48; 62pp; English.

The present sequence represents a protein encoded by open reading frame (ORF) D of a novel polynucleotide sequence designated "GS". GS is a pathogenicity island of 8 kb of DNA comprising a core region of 5.75 kb with multiple ORFs and an adjacent transmissible element of 2.5 kb. The ORFs, and also the transmissible element, encode proteins which may be linked to pathogenesis, such as providing receptors for cellular recognition. GS was discovered and characterised using differential DNA analysis technology. It is found within Mycobacterium paratuberculosis and it has also been identified in Mycobacterium avium subspecies PT silvaticum. These pathogenic mycobacteria cause chronic inflammation of the intestine and Crohn's disease in humans. The protein products of the ORFs of GS can be used for detecting mycobacteria or for diagnosing, treating or preventing mycobacterial diseases. In particular they can be used as vaccines for inflammatory diseases such as Crohn's disease or sarcoidosis in humans or John's disease in animals.

Sequence 339 AA;

Query Match 59.0%; Score 971; DB 18; Length 339;
Best Local Similarity 59.2%; Pred. No. 1.7e-90;
Matches 181; Conservative 50; Mismatches 75; Indels 0; Gaps 0;

QY 3 DKSAKIFVAGHGLVGSAIVRKLEQGFTNLVLKTHAELDLTRQADVSEFFSQEPVVYI 62
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 28 DRATPVYIAGHGLVGSAIVRFEAEAGFTNLI VRSRDEIDLTDRAATDFVSETRPQVII 87

QY 63 LAARKVGGIHANNTPADFIGVNLOIQTNVIHSAYEHGVKKLLFLGSSCIYPKEAPQPIR 122
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 88 DAARVGGIVANNTPADFI SENLRITOTLLDAAVAVRVPRLFLGSSCIYPKYAPQPPIH 147

QY 123 ESALLTASLEPTNEWYAIAKIAGIKTCQARYIQHGWDALSGMPTNLYGPNDFHPENSHV 182
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 148 ESALLTGPLETNDVAIAKIAGILQVAVRQYGLAWISAMPTNLYGPDNFSPGSHL 207

QY 183 LPALMRFRHEAKVNGAEVVWGTGSPLEFLHVDDLADACVPLLDYSLGHVNIQSQQ 242
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 208 LPALIRRYEAKAGGAEEVTNWGTGTPRRRELLHVDDLASACLFLLEHFDGNHNVTGV 267

QY 243 EVTIRELAEVLKVEVFEGKLGDCTKPDGTPRKLMDSKLASLGHTPKVSLRDGLSQTYY 302
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 268 DHSISEIADMVATAVGIGETRWDPDKPDGTPRKLDDVSALRELGLWRPRIALKOGIDATV 327

QY 303 DWYLKN 308
|| ||
DB 328 SWRTN 333

RESULT 9
AAAY81360 ID AAY81360 standard; protein; 317 AA.
XX AC AAY81360;
XX XX
DT 19-JUN-2000 (first entry)
XX DE Escherichia coli GDP-fucose synthetase.
XX KW GDP-fucose synthetase; GFS; X-ray crystallography;
KW structural determination; dual function; epimerisation;
KW NADPH-dependent reduction; agonist; antagonist; identification;
KW rational drug design; immunosuppressant; cell adhesion inhibitor;
KW selectin-mediated.
XX OS Escherichia coli.
XX PN WO200009744-A1.
XX PD 24-FEB-2000.
XX PF 13-AUG-1999; 99WO-US18441.
XX PR 13-AUG-1998; 98US-0096452.
XX PA (AMHP) AMERICAN HOME PROD CORP.
XX PI Somers WS, Stahl ML, Sullivan FX;
XX WPI; 2000-224366/19.
XX New crystalline Guanosine diphosphate-fucose synthetase for identifying agonist and antagonist of its human homolog and for the design of immunosuppressants that acts by blocking selectin mediated cell adhesion -
PS Disclosure; Fig 4; 56pp; English.
CC The invention relates to a crystalline GDP-fucose synthetase (GFS). This was used to determine the crystal structure of GFS from Escherichia coli (AAY81360) to a resolution of 2.2 Angstroms. The structure of the enzyme when bound to its cofactor NADPH and product NADP+ has also been determined. The structure of GFS revealed it to be a member of the short chain dehydrogenase/reductase (SDR) family of enzymes which have a conserved core tertiary protein fold and a serine, tyrosine and lysine catalytic triad. GFS is particularly closely related to E. coli UDP-galactose-4-epimerase (GalE, AAY81359). E. coli GFS has a human homologue, referred to as FX protein (AAY81361), which has 29% amino acid identity. GDP-fucose is synthesised in a three-step reaction from GDP-mannose, with the second and third steps (epimerisation and NADPH-dependent reduction) being catalysed by the dual function GFS to form fucose-GDP. This is then added to glycoconjugates by specific transferases that use GDP-fucose as the sugar donor. In humans, a defect in GDP-fucose biosynthesis is responsible for the immune disorder leukocyte adhesion deficiency type II. The crystalline structure of GFS derived from E. coli is useful for identifying its agonists and antagonists and is useful for the design of inhibitors of the human GFS which ultimately could lead to the design of immunosuppressants that act by blocking selectin mediated cell adhesion. The present sequence represents E. coli GFS.
SQ Sequence 317 AA;

Query Match 57.3%; Score 943.5; DB 21; Length 317;
Best Local Similarity 57.6%; Pred. No. 9.8e-88;
Matches 179; Conservative 56; Mismatches 67; Indels 9; Gaps 1

QY 7 KIFVAGHGLVGSAIVRKLEQGFTNLVLKTHAELDLTRQADVSEFFSQEPVVIIAAA 66
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 3 RVFIAGHRGWGSAIRQLQEGRGVDVELVTRDELNLDLSRAVHDFFASERIDQVYIAAA 62

QY 67 KVGGIHANNTPADFIGVNLOIQTNVIHSAYEHGVKKLLFLGSSCIYPKEAPQIPESAL 122
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 63 KVGGIVANNTPADFIYQNMMIESNIIHAHQNDVNKLLFLGSSCIYPKLAKQPMASESEL 122

QY 127 LTASLEPTNEWYAIAKIAGIKTCQARYIQHGWDALSGMPTNLYGPNDFHPENSHVLPAL 182
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 123 LQGTLEPTNEPYIAKIAGIKLCESYNRQYGRDYRVSMPTNLYGFHDFHPNSHVIPAL 182

QY 187 MRREFHAKVNGAEVVWGTGSPLEFLHVDDLADAC-----VFLLDRYSGLSHVN 233
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 183 LRRFHETAQANPADPVVWGTGTPMRRELFHVDDMAAASHVMELAHEVWLENQTQMLSHIN 242

QY 238 IGSGQEVITRELAEVLKVEVFEGKLGDCTKPDGTPRKLMDSKLASLGHTPKVSLRDG 299
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 243 VGTGVDCTIRELAQTIKAVGVKGRVVDASKPDGTPRKLDDVTRLQLGHWYHISLEAG 302

QY 298 LSQTYDWYLKN 308
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 303 LASTYQWFLEN 313

RESULT 10
AAAY29659 ID AAY29659 standard; protein; 321 AA.

XX AC AAY29659;
XX 03-NOV-1999 (first entry)
XX Escherichia coli YEF B amino acid sequence.
XX Escherichia coli; YEF B; expression vector; enzymatic conversion;
XX guanosine diphosphate-4-keto-6-deoxymannose; GDP-mannose; GDP-fucose;
XX fucosylated oligosaccharides; enzymatic fucosyl transfer; glycoprotein;
XX glycolipid; insulin; growth hormone; tissue plasminogen activator;
XX interleukin; viral antigen; therapeutic; diagnosis; food.
XX Escherichia coli.
XX WO9936555-A1.
XX 22-JUL-1999.
XX 15-JAN-1999; 99WO-US00893.
XX 14-JAN-1999; 99US-0071076.
XX 15-JAN-1998; 98US-0071076.
XX (CYTE-) CYTEL CORP.
XX Sjoberg ER;
XX WPI; 1999-527202/44.
XX New vector expressing an enzyme that converts guanosine
XX diphosphate-4-keto-6-deoxymannose to GDP-fucose, used to prepare
XX fucosylated oligosaccharides
XX Claim 41; Page 77; 79pp; English.
XX The present invention describes an expression vector comprising a
XX promoter linked to a nucleic acid (I) that encodes a prokaryotic enzyme
XX (II) having both epimerase and reductase activity, for the catalysis of
XX conversion of GDP (guanosine diphosphate)-4-keto-6-deoxymannose (III) to
XX GDP-fucose (IV), where the vector lacks an Escherichia coli wcaI coding
XX region. An expression vector is specifically claimed where (I) encodes
XX Escherichia coli YEF B. The present sequence represents the E. coli
XX YEF B protein. (II) is used for production of (IV) which is then used to
XX prepare fucosylated oligosaccharides (A) by enzymatic fucosyl transfer,
XX e.g. to modify oligosaccharide components of glycoproteins or
XX glycolipids, such as insulin, human or bovine growth hormones, tissue
XX plasminogen activator, interleukins and viral antigens, or chimeric
XX products such as immunoglobulin derivatives. (A) are variously useful as
XX therapeutic and diagnostic agents and in foods. Combining two activities
XX in a single enzyme simplifies the process, allowing efficient synthesis
XX of complex fucosylated oligosaccharides in a single reaction vessel from
XX readily available starting materials. The method is suitable for large
XX scale synthesis, e.g. 0.2 kg batches. (II) can be expressed efficiently
XX in prokaryotic cells (contrast similar mammalian enzymes).
XX Sequence 321 AA;
Query Match 57.0%; Score 937.5; DB 20; Length 321;
Best Local Similarity 56.3%; Pred. No. 4.1e-87;
Matches 177; Conservative 58; Mismatches 67; Indels 9; Gaps 1;
QY 7 KIFVAGHRLVGSIAVRKLEQCFNLVLKTHAELDLTRQADVESPFSEKPVYVILAAA 66
DB 5 RVFIAGHGMVGSIAIRRLEQGVDELVRDELNLLDSRAVHDFPAGERIDQVYVILAAA 64
QY 67 KVGGIHANNTPADFGVNLQITNVVHSAYEHGKVLFLGSSCIYKPFAPQIPESAL 126
DB 65 KVGGIHANNTPADFGVNNMTIESNIIHAHQNDVNNKLLFLGSSCIYKPLAQPMASEL 124
QY 127 LTASLEPTNEWAIAGIKTKCOAYRIOGHWDASGMFTNLYGPNDFNPENSHVLPAL 186
DB 125 LGTLEPTNEPVAIAKIAGIKLCSYRQYGRDYRSWMTNLYGPNDFNPENSHVLPAL 184

QY 187 MRRFHEAKVNGAEVVMVGTGSPLEFLHVDLADAC-----VFLLDRYSGLHVN 237
DB 185 LRRFHEATAQNAAPDVVMVGTGSPLEFLHVDLADAC-----VFLLDRYSGLHVN 244
QY 238 IGSQGEVITRELAEVLKVEVVGEGKLGWDCTKPGDTPRKLMDSKSLASLGWTPKVSRLRG 297
DB 245 VGTGVDCITRDVAQTIKAVGVYKGRWFDASKPDGTPRKLLDVTRLHQLGMYHEISLEAG 304
QY 298 LSQTYDWMYLN 308
DB 305 LASTYQWFLEN 315
RESULT 11
AAY54115
ID AAY54115 standard; Protein; 321 AA.
XX AAY54115;
XX 27-MAR-2000 (first entry)
XX A GDP-4-keto-6-deoxy-D-mannose epimerase/reductase protein.
XX GDP-4-keto-6-deoxy-D-mannose epimerase/reductase; GDP-D-mannose;
XX GDP-L-galactose; vitamin C; ascorbic acid; L-ascorbic acid;
XX ascorbic acid pathway enzyme; hexokinase; Glucose phosphate isomerase;
XX phosphomannose isomerase; phosphomannomutase; L-galactose dehydrogenase;
XX GDP-D-mannose pyrophosphorylase; GDP-D-mannose:GDP-L-galactose epimerase;
XX GDP-L-galactose phosphorylase; L-galactose-1-P-phosphatase;
XX L-galactono-gamma-lactone dehydrogenase; ester.
XX Escherichia coli.
XX WO9964618-A1.
XX 16-DEC-1999.
XX 26-MAY-1999; 99WO-US11576.
XX 08-JUN-1998; 98US-0088549.
XX 17-MAR-1999; 99US-0125073.
XX 18-MAR-1999; 99US-0125054.
XX (DCVB-) DCV INC DBA BIO-TECH RESOURCES.
XX Berry A, Running JA, Severson DK, Burlingame RP;
XX WPI; 2000-105890/09.
XX N-PSDB; AA245316.
XX Production of ascorbic acid or esters, using microorganisms or plants
XX which have genetic modification in enzymes involved in the ascorbic
XX acid synthesis pathway -
XX Claim 27; Page 170-171; 187pp; English.
XX The present sequence represents a GDP-4-keto-6-deoxy-D-mannose epimerase/
XX reductase. The enzyme catalyses the conversion of GDP-D-mannose to
XX GDP-L-galactose. The enzyme can be modified, and used to produce
XX transgenic microorganisms, which can be used in fermentation techniques
XX to produce vitamin C (ascorbic acid, L-ascorbic acid). The enzyme is
XX modified to increase its action. Other ascorbic acid pathway enzymes
XX which may be used in the method of the invention include hexokinases,
XX glucose phosphate isomerases, phosphomannose isomerases,
XX phosphomannomutases, GDP-D-mannose pyrophosphorylases,
XX GDP-D-mannose:GDP-L-galactose epimerases, GDP-L-galactose phosphorylases,
XX L-galactose-1-P-phosphatases, L-galactose dehydrogenases, and
XX L-galactono-gamma-lactone dehydrogenases. The methods can be used for
XX the production of ascorbic acid or esters using microorganisms or plants.
XX Sequence 321 AA;

Query Match 57.0%; Score 937.5; DB 21; Length 321;
 Best Local Similarity 56.9%; Pred. No. 4.1e-87;
 Matches 177; Conservative 58; Mismatches 67; Indels 9; Gaps 1;

QY 7 KIFVAGHRLGVSATVRKLEQGTNVLKTHAELDLTRQADVESFSEKPKVYVILAAA 66
 DB 5 RVFIAGHGMVGSATRRLEQGDVELVLRTRDELNLDSRAVDFPASERIDQVYLA 64

QY 67 KVGGIHANNYPADFIQVNLQIQTNNVIHSAIEHGVKLLFLGSSCIYKPKFAPQIPESAL 126
 DB 65 KVGGIHANNYPADFIQVNLQIQTNNVIHSAIEHGVKLLFLGSSCIYKPKFAPQIPESAL 124

QY 127 LTASLEPTNEWAIKIAIKTCQAYRIQHGWDALSGMPTNLYGPNDFHPNSHVLPAL 186
 DB 125 LQGLEPTNEPVAIAIKIAGIKLCESYNRQGRDYRSVMTNLYGPNDFHPNSHVLPAL 184

QY 187 MRRFHEAKVNGAEVVGWGTGSPLEFLVDDLDAC-----VFLLDRYSGLEHVN 237
 DB 185 LRRFHEATQANPDVVVWGSGTPMRFLVHVDVDAASIHVMELAEHVEWLENTQPMLSHIN 244

QY 238 IGSQGOEVTIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSKSLASLGWTPKVSRLDG 297
 DB 245 VGTGVDCITRDVAQITAKVVGKGRVFDASKPDGTPRKLLDVTRLHQLGWYHEISLEAG 304

QY 298 LSQTYDWYLN 308
 DB 305 LASTYQWFLN 315

RESULT 12
 AAW88319 standard; Protein; 323 AA.

QY AAW88319;
 DB AAW88319;

QY 26-APR-1999 (first entry)
 DB 26-APR-1999 (first entry)

QY E. coli O157 antigen pathway ORF8 protein product.
 DB E. coli O157 antigen; diarrhoea; haemorrhagic colitis;

QY O antigen; O157 antigen; diarrhoea; haemorrhagic colitis;
 DB O antigen; O157 antigen; diarrhoea; haemorrhagic colitis;

QY Escherichia coli.
 DB Escherichia coli.

QY WO9850531-A1.
 DB WO9850531-A1.

QY 12-NOV-1998.
 DB 12-NOV-1998.

QY 01-MAY-1998; 98WO-AU00315.
 DB 01-MAY-1998; 98WO-AU00315.

QY 22-JUL-1997; 97AU-0008162.
 DB 22-JUL-1997; 97AU-0008162.

QY 01-MAY-1997; 97AU-0006545.
 DB 01-MAY-1997; 97AU-0006545.

QY (UNSY) UNIV SYDNEY.
 DB (UNSY) UNIV SYDNEY.

QY Reeves PR, Wang L;
 DB Reeves PR, Wang L;

QY WPI; 1999-059669/05.
 DB WPI; 1999-059669/05.

QY N-PSDB; AAX06749.
 DB N-PSDB; AAX06749.

QY Nucleic acid molecules specific for bacterial polysaccharide
 PT antigens - useful for detecting specific strains in, e.g. food,
 PT faeces or patient samples

QY Disclosure; Fig 8; 165pp; English.
 DB Disclosure; Fig 8; 165pp; English.

QY This is the amino acid sequence of a protein encoded by an open
 CC reading frame (ORF8) of a gene cluster (see AAX06749) involved in
 CC the biosynthesis of the Escherichia coli O157 O antigen. The
 CC protein shows homology to the E. coli Wcag colanic acid capsule
 CC protein. The use of nucleic acids derived from assembly and
 CC transport genes, particularly wbd (transferase), wzx (flippase)
 CC and wzy (polymerase) genes, within O antigen gene clusters

CC improves the specificity of methods for the detection and
 CC identification of O antigens, e.g. in testing food- or faecal-
 CC derived samples, or samples from patients. The O antigen is a
 CC major virulence factor of enteropathogenic E. coli strains that
 CC cause diarrhoea and haemorrhagic colitis.

QY Sequence 323 AA;
 DB Sequence 323 AA;

QY Query Match 55.6%; Score 915.5; DB 20; Length 323;
 DB Best Local Similarity 57.2%; Pred. No. 7.4e-85;
 DB Matches 178; Conservative 50; Mismatches 74; Indels 9; Gaps 2;

QY 7 KIFVAGHRLGVSATVRKLEQGTNVLKTHAELDLTRQADVESFSEKPKVYVILAAA 66
 DB 7 RFIAGHGMVGSATRRLEQGDVELVLRTRDELNLDSRAVDFPASERIDQVYLA 66

QY 67 KVGGIHANNYPADFIQVNLQIQTNNVIHSAIEHGVKLLFLGSSCIYKPKFAPQIPESAL 126
 DB 67 KVGGIHANNYPADFIQVNLQIQTNNVIHSAIEHGVKLLFLGSSCIYKPKFAPQIPESAL 126

QY 127 LTASLEPTNEWAIKIAIKTCQAYRIQHGWDALSGMPTNLYGPNDFHPNSHVLPAL 186
 DB 127 LQGLEPTNEPVAIAIKIAGIKLCESYNRQGRDYRSVMTNLYGPNDFHPNSHVLPAL 186

QY 187 MRRFHEAKVNGAEVVGWGTGSPLEFLVDDLDACVFLDR-Y-----GLEHVN 237
 DB 187 LRRFHEAVENNSPNVVGSGTPKREFLVHVDVDAASIHVMELAEHVEWLENTQPMLSHIN 246

QY 238 IGSQGOEVTIRELAELVKEVVGEGKLGWCTKPDGTPRKLMDSKSLASLGWTPKVSRLDG 297
 DB 247 IGTGIDCTICEAETIAKVVGVKGHITFTDTPDGAPRKLLDVTLLHQLGWNHKTILHKG 306

QY 298 LSQTYDWYLN 308
 DB 307 LENTYQWFLN 317

RESULT 13
 ABG18368 standard; Protein; 576 AA.

QY ABG18368;
 DB ABG18368;

QY 18-FEB-2002 (first entry)
 DB 18-FEB-2002 (first entry)

QY Novel human diagnostic protein #18359.
 DB Novel human diagnostic protein #18359.

QY Human; chromosome mapping; gene mapping; gene therapy; forensic;
 DB Human; chromosome mapping; gene mapping; gene therapy; forensic;
 DB food supplement; medical imaging; diagnostic; genetic disorder.

QY Homo sapiens.
 DB Homo sapiens.

QY WO200175067-A2.
 DB WO200175067-A2.

QY 11-OCT-2001.
 DB 11-OCT-2001.

QY 30-MAR-2001; 2001WO-US08631.
 DB 30-MAR-2001; 2001WO-US08631.

QY 31-MAR-2000; 2000US-0540217.
 DB 31-MAR-2000; 2000US-0540217.

QY 23-AUG-2000; 2000US-0649167.
 DB 23-AUG-2000; 2000US-0649167.

QY (HYSE-) HYSEQ INC.
 DB (HYSE-) HYSEQ INC.

QY Drmanac RT, Liu C, Tang YT;
 DB Drmanac RT, Liu C, Tang YT;

QY WPI; 2001-639362/73.
 DB WPI; 2001-639362/73.

QY N-PSDB; AAS82555.
 DB N-PSDB; AAS82555.

QY New isolated polynucleotide and encoded polypeptides, useful in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits and to assess
 CC biodiversity -

PS Claim 20; SEQ ID NO 48727; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 576 AA;

Query Match 43.7%; Score 718.5; DB 22; Length 576;

Best Local Similarity 58.2%; Pred. No. 2.3e-64; Indels 9; Gaps 1;

Matches 142; Conservative 38; Mismatches 55;

QY 23 RKLQEGFTNLVLKTHAELDLTRQADVESFFSOKEPVYVILAAKVGGIHANNYPADFI 82

Db 293 RQLQEGDVELVLRDELNLLDSRAVHDFFAFERIDQVYLAARKVGIVANNYPADFI 352

QY 83 GVNLIQTQNVHSAHGVKLLFLGSSCTYPKFAFQPIPEASALLTASLEPTNEWVIAIK 142

Db 353 YQNMIESIIHAAHQNDVKNLLFLGSSCTYPKFAFQPIPEASALLTASLEPTNEWVIAIK 412

QY 143 IAGIKTCQAVRIQHGWDATSGMPTNLYGPNDFHPSNHLVLPALRRFHEAKVNGAEVYV 202

Db 413 IAGIKLCEYNQYGRDYRSVMPTNLYGPHDNFPHSNHVPALLRRFHEATAQNAPOVV 472

QY 203 VMGTGSPFLREFLHVDLADAC-----VFLLDRYSGLEHVNIGSGQEVITIRELAELV 253

Db 473 VMGSGTPMREFLHVDLADAC-----IHMELAEHVEWLENTPMLSHINVGTGVDCTIRELAQTI 532

QY 254 KEVV 257

Db 533 GKVV 536

RESULT 14

AAW00029

ID AAW00029 standard; Protein; 162 AA.

XX AAW00029;

XX 28-SEP-2001 (first entry)

XX GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase-4-reductase cDNA sequence #97.

XX Moss; carbohydrate metabolism related protein; CWRP; sugar; cofactor;

KW fine chemical production; carbohydrate; polysaccharide.

XX Physcomitrella patens.

OS WO200144476-A2.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000NO-EP12697.

XX 16-DEC-1999; 99US-0171101.

XX (BADI) BASF PLANT SCI GMBH.

XX Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;

PI Frank M, Freund A, Duwenig E, Schmidt R, Reski R;

XX WPI; 2001-398155/42.

DR N-PSDB; AAH88715.

XX Novel moss nucleic acid molecules encoding a carbohydrate metabolism

PT related protein useful for modulating production of fine chemicals such

PT as carbohydrates, cofactors and enzymes from microorganisms and plants

XX Claim 30; Page 114; 133pp; English.

XX This invention relates to nucleic acid molecules AAH88708 - AAH88796

CC isolated from Physcomitrella patens (a moss), which encode carbohydrate

CC metabolism related proteins (CWRP) represented in AAM00022 - AAM00110.

CC Included in the invention is a vector containing the CWRP cDNA, and a

CC host cell transformed with the vector. The host cell (a microorganism,

CC Corynebacterium or Brevibacterium, moss or algae or a plant cell) is

CC useful for producing a fine chemical such as carbohydrates, cofactors

CC and/or enzymes. The nucleic acid molecules are suitable for modifying a

CC carbohydrate production system in a host, e.g., microorganisms and

CC plants. They are also useful to identify those DNA sequences and enzymes

CC in other species which are useful to modify the biosynthesis of starch,

CC utilised in the genetic engineering of Corynebacterium glutamicum and the

CC related Brevibacterium species and Acetobacter xylinum and Chlorella to

CC make it a better or more efficient producer of one or more fine

CC chemicals. Mutagenesis of one or more CWRPs may also result in CWRPs

CC having altered activities which indirectly impact the production of one

CC or more desired fine chemicals from plants. Primers AAH88705 - AAH88707

CC are used in the sequencing of the CWRP cDNA sequences of the invention.

XX Sequence 162 AA;

Query Match 40.8%; Score 671; DB 22; Length 162;

Best Local Similarity 74.8%; Pred. No. 2.7e-60;

Matches 119; Conservative 25; Mismatches 15; Indels 0; Gaps 0;

QY 150 QAYRIQHGWDAISGMTNLYGPNDFHPSNHLVLPALRRFHEAKVNGAEVYVWGTGSP 209

Db 1 QAYRIQYNFADISGMTNLYGPHDNFPHPSNHLVLPALRRFHEAKVNGAEVYVWGTGSP 60

QY 210 LREFLHVDLADACVFLLDYRSGLEHVNIGSGQEVITIRELAELVKEVVGPEGLGWDCTK 269

Db 61 FREFLHVDLAEATVFLQLQYSAHEHVNMGSGSEVSIKELAEVKEVVGFGQLTWDTSK 120

QY 270 PDGTPRKLMDSSKLASLGWTPKVSLRDCLSQTYDHYLKN 308

Db 121 PDGTPRKLIDSSKLANMGWQARIPLKEGLAETRYKWCEN 159

RESULT 15

AAW88305

ID AAW88305 standard; Protein; 307 AA.

XX AAW88305;

XX 26-APR-1999 (first entry)

XX E. coli O111 antigen gene cluster ORF6 (wbdJ) protein product.

XX O antigen; O111 antigen; wbdJ gene; diarrhoea;

KW haemorrhagic colitis; diagnosis.

XX Escherichia coli.

XX WO980531-A1.

XX 12-NOV-1998.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2003, 02:39:05 ; Search time 1673 Seconds
(without alignments)

9060.954 Million cell updates/sec

Title: US-10-089-014-2

Perfect score: 936

Sequence: 1 atgtctgacaaatcgccaa.....tgaagaatgtttgcaaccca 936

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	527	56.3	627	9 AU237351	AU237351 AU237351
C 2	513	54.8	537	17 CNS00P9U	AL084240 Arabidops
C 3	436.2	46.6	441	17 B22510	B22510 F20D2TF IGF
4	417.8	44.6	856	12 BF276275	BF276275 GA_Eb002
5	416.6	44.5	852	12 BF278112	BF278112 GA_Eb003
C 6	408.2	43.6	1439	11 AY105324	AY105324 Zea mays

7	386	41.2	618	17	AQ841772
8	354.8	37.9	652	10	AW041037
9	354.4	37.9	731	17	BH173899
10	352.2	37.6	566	17	BH173840
11	350.6	37.5	752	14	BQ116267
12	339.8	36.3	687	12	BG440809
13	338.6	36.2	614	9	AI897511
14	326	34.8	664	9	AI726948
15	322.2	34.4	590	9	AI487671
16	314.8	33.6	539	13	BI424065
17	308.6	33.0	547	10	BE458430
18	308.6	33.0	665	14	BQ405379
19	299.8	32.0	649	10	AW761509
20	295.2	31.5	739	10	BE055323
21	295	31.5	513	13	BI498273
22	290.8	31.0	497	10	AW507622
23	290.2	31.0	512	10	BE459038
24	287.8	30.7	748	17	BH547992
C 25	282	30.1	667	14	BQ411981
26	277.8	29.7	547	10	AW906983
27	268.8	28.7	302	14	Z35041
28	265.4	28.4	539	13	BM110048
29	258.6	27.6	529	10	BE343057
30	256.6	27.4	737	9	AF109667
C 31	256.4	27.4	890	17	BI9210
32	254	27.1	522	9	AI726475
33	253.6	27.1	757	10	BE660172
34	250.4	26.8	481	13	BI433478
35	246.8	26.4	603	13	BI893509
36	241.2	25.8	669	14	BM779643
37	240.6	25.7	472	10	BE463183
38	240	25.6	396	12	BQ405837
C 39	239.6	25.6	648	12	BE823790
C 40	235.8	25.2	239	17	BH851675
41	230.2	24.6	541	13	BI699441
42	230	24.6	622	10	BE440335
43	229.6	24.5	570	13	BM371263
44	228.6	24.4	657	14	BQ408320
C 45	227	24.3	437	10	AV560875

ALIGNMENTS

RESULT 1

AU237351

LOCUS AU237351

DEFINITION AU237351

VERSION AU237351

KEYWORDS EST.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 627)

AUTHORS Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shingawa, A., Muramatsu, M., Hayashizaki, Y., and Shinozaki, K.

TITLE Large scale analysis of Arabidopsis full-length cDNA

JOURNAL Unpublished (2002)

COMMENT Contact: Motoaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: mseki@tc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

627 bp mRNA linear EST 01-APR-2002
AU237351 RAFL16 Arabidopsis thaliana cDNA clone RAFL16-16-116 5',

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

EST 01-APR-2002

and XhoI was ligated to modified Lambda FUC-1 vector (Carinci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified paluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES	source	Location/Qualifiers
		1. 627
		/organism="Arabidopsis thaliana"
		/db_xref="taxon:3702"
		/clone="RAFL16-16-116"
		/clone_lib="RAFL16"
		/lab_host="DH10B"
		/note="Site 1: BamHI; Site 2: SalI; dark-grown"
BASE COUNT	151 a	159 c 125 g 189 t 3 others

JOURNAL REFERENCE	Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
AUTHORS	Unpublished
TITLE	2 (bases 1 to 537)
JOURNAL	Genoscope.
FEATURES	Direct Submission
source	Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr) - Web : www.genoscope.cns.fr) Location/Qualifiers 1. 537 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /clone="F8H3" /clone_lib="IGF" /notes="end : T7"
BASE COUNT	164 a 110 c 131 g 132 t
ORIGIN	
Query Match	54.8%; Score 513; DB 17; Length 537;
Best Local Similarity	99.8%; Pred. No. 3.9e-140;
Matches	524; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1	ATGTCACAAAATCTGCCAAAATCTTCGTCGCGGGTCATCGTGGTTTGGTTGGATCTGCC 60
DB 524	ATGTCACAAAATCTGCCAAAATCTTCGTCGCGGGTCATCGTGGTTTGGTTGGATCTGCC 465
QY 61	ATTGTCGCGAAGCTTCAGGAACAAGGTTTCACCAATCTCGTCTCTAAAAACACACGCCGAG 120
DB 464	ATTGTCGCGAAGCTTCAGGAACAAGGTTTCACCAATCTCGTCTCTAAAAACACACGCCGAG 405
QY 121	CTTGATCTCACTCGTCAAGCCGATGTGAATCTCTTTTCTCAAGAGAGCCAGTTTAT 180
DB 404	CTTGATCTCACTCGTCAAGCCGATGTGAATCTCTTTTCTCAAGAGAGCCAGTTTAT 345
QY 181	GTAATCCTAGCAGCCTAAAGTTGCTGGTATTTCAGCTTAAACAACCTATCTCGCTGAT 240
DB 344	GTAATCCTAGCAGCCTAAAGTTGCTGGTATTTCAGCTTAAACAACCTATCTCGCTGAT 285
QY 241	TTCAATGGTGTCAATCTCCAGATTTCAGACCAATGTGATCCACTCTGCATATGAGCACGGT 300
DB 284	TTCAATGGTGTCAATCTCCAGATTTCAGACCAATGTGATCCACTCTGCATATGAGCACGGT 225
QY 301	GTGAAGAAGCTTCTCTTCCTTGGATCATCTGATTTTACCTAAATTTGCTCTCAGCCA 360
DB 224	GTGAAGAAGCTTCTCTTCCTTGGATCATCTGATTTTACCTAAATTTGCTCTCAGCCA 165
QY 361	ATTCTCTGAGTCCTTTGTTTAAACAGATCGCTTGAACCAACTAATCAGTGGTATGCTATT 420
DB 164	ATTCTCTGAGTCCTTTGTTTAAACAGATCGCTTGAACCAACTAATGAGTGGTATGCTATT 105
QY 421	GCTAAGATCGCTGGGATTAAAGCTTCTCAGGCTTTATAGGATTCAGCACGGATGGATGCA 480
DB 104	GCTAAGATCGCTGGATTAAAGCTTCTCAGGCTTTATAGGATTCAGCACGGATGGATGCA 45
QY 481	ATCTCTGGCAGTGCCTACTAATCTCTATGGTCTCTAATGACAAATTC 525
DB 44	ATCTCTGGCAGTGCCTACTAATCTCTAT -GTCCTAATGACAAATTC 1

AUTHORS Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Coffey, E., Golden, K., Johnson, K., Adams, M.D. and Venter, J.C.
TITLE A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: F20D2TR
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13-21
Class: BAC ends
High quality sequence stop: 441.
Location/Qualifiers
1. 441
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F20D2"
/clone_lib="IGF"
/sex="Hermaphrodite"
/note="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"
BASE COUNT 133 a 87 c 108 g 113 t
ORIGIN

Query Match 46.6%; Score 436.2; DB 17; Length 441;
Best Local Similarity 99.3%; Pred. No. 1.6e-117;
Matches 438; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 94 AATCTCGTTCTTAAACACACGCCGAGCTTGATCTCACTCGTCAGCCGATGTGAATCC 153
DB |||||
441 AATCTAGTTCTTAAACACACGCCGAGCTTGATCTCACTCGTCAGCCGATGTGAATCC 382
QY 154 TTCTTTTCTCAAGAGAAGCCAGTTTATGTAATCTTAGCAGCAGCTAAAGTTGGTGTATT 213
DB |||||
381 TTCTTTTCTCAAGAGAAGCCAGTTTATGTAATCTTAGCAGCAGCTAAAGTTGGTGTATT 322
QY 214 CAGCTTAACACACCTATCTCTGATTTCAATGTTGTTGTTCAATCTCCAGATTGAGCAAT 273
DB |||||
321 CAGCTTAACACACCTATCTCTGATTTCAATGTTGTTGTTCAATCTCCAGATTGAGCAAT 262
QY 274 GTGATCCACTCTGCATATGAGCAGCGTGTGAAGAGCTTCTCTCTTGGATCATCTGC 333
DB |||||
261 GTGATCCACTCTGCATATGAGCAGCGTGTGAAGAGCTTCTCTCTTGGATCATCTGC 202
QY 334 ATTTACCCCTAAATTTGCTCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCATCGCTT 393
DB |||||
201 ATTTACCCCTAAATTTGCTCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCATCGCTT 142
QY 394 GAACCAACTAATGAGTGTATGCTATTGTTAGATCGCTGGGATTAAGCTTGTAGCT 453
DB |||||
141 GAACCAACTAATGAGTGTATGCTATTGTTAGATCGCTGGGATTAAGCTTGTAGCT 82
QY 454 TATAGGATTCAGCAGCGATGGGATGCAATCTCTGGCATCGCTACTAATCTCTATGCTCT 513
DB |||||
81 TATAGGATTCAGCAGCGATGGGATGCAATCTCTGGCATCGCTACTAATCTCTATGCTCT 22
QY 514 AATGACAATTTCCACCCGGAG 534
DB |||||
21 AATGACAATTTCCACCCGGAG 1

RESULT 4
BF276275
LOCUS BF276275 856 bp mRNA linear EST 07-MAR-2001
DEFINITION GA_Eb0028C16f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Eb0028C16f, mRNA sequence.
ACCESSION BF276275
VERSION BF276275.1 GI:11207345

KEYWORDS Gossypium arboreum.
SOURCE Gossypium arboreum
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 856)
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE An integrated analysis of the genetics, development, and evolution
of the cotton fiber
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCATATAGG
High quality sequence stop: 712.
Location/Qualifiers
1. 856
/organism="Gossypium arboreum"
/strain="ARA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Eb0028C16f"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 213 a 217 c 208 g 217 t 1 others
ORIGIN

Query Match 44.6%; Score 417.8; DB 12; Length 856;
Best Local Similarity 70.2%; Pred. No. 6.4e-112;
Matches 560; Conservative 0; Mismatches 238; Indels 0; Gaps 0;
QY 2 TCTCTGACAAATCTGCCAAATCTTCGTCGGGTCATCGTGTGGTTCGATCGCA 61
DB |||||
51 TCTCGAAACAAATCAGCGAAGATCTTCGTAGCAGGCGATCGAGGCTTAGTGTTCGCCA 110
QY 62 TTGTCCGCAAGCTTCAGGAACAAGGTTTCAACAAATCTGTTCTTAAACAACACGCGAGC 121
DB |||||
111 TAGTTCGTAAATTCAGTCCCTAGCTTTACCAACCTCTACTCCGACCCATCGCATC 170
QY 122 TTGATCTCACTGCTCAAGCCGATGTAATCTCTTTTCTCAAGAGAACGAGTTATG 181
DB |||||
171 TGGACCTTACTCGCCAATCCGACGTGCAATCTCTTCGCCGATGAGAAACCTCACTATG 230
QY 182 TAATCTAGCAGCAGCTAAAGTTGGTGTATTCAGCTAACCAACCTATCTGCTGATT 241
DB |||||
231 TGTACTAGTCCCGCTAAAGTTGGTGGATCCAGCCAACCACTTACCCCTGCCGATT 290
QY 242 TCATTGGTGTCAATCTCCAGATTGAGCAATGTGATCCACTCTCTGATATGACGCGTG 301
DB |||||
291 TCATTGCCATCACTCCAAATCCAGACCAAGTCAATGATTCCTTACCCGCCGCGG 350
QY 302 TGAAGAAGCTTCTCTTCTTGGATCATCTGCAATTTACCCATAATTTGCTCTCAGCCAA 361
DB |||||
351 TTAAGAAAATCTCTTCTCGGTTCTCTTGTATTATCCCAAAATTTGGCGGCAACCCA 410
QY 362 TTCCTGAGTCTGTTTGTATACAGCATCGCTTGAACCACTAATGAGTGTATGCTATTG 421
DB |||||
411 TCCCTGAACACGCGCTCTTATCCGTCCTCCCTGGAACCCCAACCAAGTGTATGCGGTTG 470
QY 422 TTAAGATCGCTGGGATTAAGACTTCTCAGGCTTATAGGATTCAGCAACGATGGGATGCAA 481
DB |||||
471 CCAAAATCGGGGATCAAAATGTCCAAAGCGTACAGATTCCAGCAGCGCTTCGATGCTA 530
QY 482 TCTCTGGCATGCTACTAATCTCTATGTTCTTAATGCAATTTCCACCCGGAGAAATTCCTC 541

FEATURES		Missouri, Columbia, MO 65211, USA	
source	Location/Qualifiers	1. .1439	
	/organism="Zea mays"		
	/db_xref="MaizeDB:638464"		
	/db_xref="taxon:4577"		
	/clone="PCO143177"		
Library		/clone_lib="Maize Mapping Project/DuPont Cornsensus	
Library		/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"	
BASE COUNT		297 a	357 c 443 g 342 t
ORIGIN			
Query Match 43.6%; Score 408.2; DB 11; Length 1439;			
Best Local Similarity 65.8%; Pred. No. 6e-109;			
Matches 609; Conservative 0; Mismatches 313; Indels 3; Gaps 1;			
QY	5	CTGACAAATCTGCCAAATCTTCGTCGGGTTCATCGTGGTTGGTGGATCTGCCATG	64
DB	1400	CCGACAAGGACCAAGGTTCTTCGTCGGCGCACCGGGCTCTGTCGGCTCGCCATCG	1341
QY	65	TCGCGAAGCTTCAGGAACAAGGTTTACCAATCTCGTTCTTAAACACACGCGGAGCTTG	124
DB	1340	TGCGCGCTCTCTCTCGTAGGTTTACCTCGTCTGTCGCGCACCCACCGCGAGCTCG	1281
QY	125	ATCTCACTCGTCAAGCCGATGTGAATCTTCTTTTCAAGAGAACCGAGTTATGTAA	184
DB	1280	ACCTGACTCGCAGCGGACGTCGAGGCTTCTTCGCGCAGAACCGCGCTACGTGG	1221
QY	185	TCCTAGCAGAGCTAAAGTTGTGTATTACGCTTAACAACACCTATCTCTGCTGATTCA	244
DB	1220	TCTTCGCGCGCCAGGTCGCGGAATCCAGCAATCTCCACCTTCCCGCGCACTTCA	1161
QY	245	TTGGTGTCAATCTCCAGATTCCAGACCAATGTGATCCACTCTGCATATGACGACGG	---TG 301
DB	1160	TGCGCGCAACCTCCAGATCCAGACCAAGTCTGCGACCGCGCTCGCTCGGCTCGG	1101
QY	302	TGAAGAAGTTCTTCTTGGATCATCTCTGATTTACCTTAAATTTGCTCTCTCAGCCAA	361
DB	1100	TCGCGAAGCTCTTCTCTCGGCTCTCTCTGATCTACCCCAAGTTTCGCGCGCAGCCA	1041
QY	362	TTCTGAGTCTGCTTTGTTAAAGCATCGCTTGAACCACTAATGATGATGATGATGATG	421
DB	1040	TACGAGGCGCGCTCTCTCTCGGCGCGCTCGAGCCCAACGAGTGTGTACGCGCTCG	981
QY	422	CTAAGATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACCGATGGATGCAA	481
DB	980	CCAAGATCGCGGCATCAAGATGTGCGAGGCTTACCGCATCCAGACGCGCTCGACGCG	921
QY	482	TCTCTGCAATGCTCTAATCTCTATGGTCTTAAATGACAATTTCCACCGGAGAAATTC	541
DB	920	TCTCTGCAATGCTCTAATCTCTATGGTCTTAAATGACAATTTCCACCGGAGAACTCG	861
QY	542	ATGTGCTTCTGCTTATGAGGAGTTTCCAGGCGGAAAGTGAATGAGCGGAGGAG	601
DB	860	ACGTCTTCTGCGGCTCATCTCGTCTTCCAGGCGCAAGGCTACTAACGCCGCCGAGG	801
QY	602	TTGTGTTGGGGTACAGGTAGTTCGTTGAGGAGTTCTTGATGTTGATGTTGGCTG	661
DB	800	TGTCGTCTGGGATCAGGCTCGCTCTGCGAGTTCTGCAAGTCTGATGATCTCGCG	741
QY	662	ATGCTTGTGTTTCTTCTGATTCGATACAGCGGTTGGAGCATGTTAAATTTGAAGTG	721
DB	740	ACGCGTCACTTCTCTGATGGATCACTACTTCGCGCATGGAGCATGTCAATGTGGGAGTG	681
QY	722	GTCAGAAGTGAATATTAGAGATGTTGGCTGAGTTGGTGAAGAGGTTGTTGTTTGAAG	781
DB	680	GGAGTGAGGTACCATCAAGGAGCTCGCCAACTAGTCAAGAGGTTGGTTCGCGCTTCCAGG	621

QY	782	GGAAGCTTGGATGGGATTCGACTAAGCCAGATGGCACACCGAGGAACTTATGGACAGCT	841
DB	620	GGAACCTGGTGTGGGACTCCAGCAAGCCGACCGCACACCCAGGAAGCTCATGGATAGCT	561
QY	842	CAAAGCTCGGCTCTTTGGTTGGACACCTAAGTTTCTTTAGATGCTCTGAGCCAAA	901
DB	560	CAAGATACAGGGATGGGTGGAAGCCCAAGATTCGCCCTCAAGGAAGCCCTCGTGAAA	501
QY	902	CTTATGATTGGTATTTTGAAGAATGT	926
DB	500	CCTACAATGGTAGTCGAGATGT	476
RESULT 7			
LOCUS AQ841772			
DEFINITION T134022 Soybean RFLP probe Glycine max genomic clone A890, DNA			
sequence.			
ACCESSION AQ841772			
VERSION AQ841772.1 GI:5930797			
KEYWORDS GSS.			
SOURCE soybean.			
ORGANISM Glycine max			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;			
Glycine.			
REFERENCE 1 (bases 1 to 618)			
AUTHORS Kim,D., Limpens,E., Peng,H., Ellis,L. and Cook,DR.			
TITLE Analysis of soybean RFLP probes			
JOURNAL Unpublished (1999)			
COMMENT Contact: Cook DR			
The Crop Biotechnology Center			
Texas A&M University			
Department of Plant Pathology and Microbiology, Rm 120 L.F.			
Peterson Bldg, College Station, TX 77843-2132, USA			
Tel: 409 845 8743			
Fax: 409 862 4790			
Email: dcook@ppserver.tamu.edu			
Other name: Soybean RFLP probe; date: 4/20/99; Submitted to the			
Database of Genome Survey Sequences (GSS) on 09/27/99; DNA sequence			
of Soybean RFLP probe A890. Clones were obtained from Biogenetic			
Services, Inc. More information is available at			
'http://chryslr.tamu.edu/medicago'.			
Seq primer: T3 (AATTAAACCTCACTAAAGG)			
Class: RFLP probe.			
FEATURES			
source			
Location/Qualifiers			
1. .618			
/organism="Glycine max"			
/db_xref="taxon:3847"			
/clone="A890"			
/clone_lib="Soybean RFLP probe"			
/note="Vector: pBS+; Site_1: PstI; Site_2: PstI"			
BASE COUNT 161 a 123 c 164 g 170 t			
ORIGIN			
Query Match 41.2%; Score 386; DB 17; Length 618;			
Best Local Similarity 76.5%; Pred. No. 1.3e-102;			
Matches 473; Conservative 0; Mismatches 145; Indels 0; Gaps 0;			
QY	201	AGTTGGTGGTATTCACGCTAACACACCTATCTCTGCTGATTTTCATTGGTGCAATCTCA	260
DB	1	AGTTGGTGGCATCCATGCCAACACACCTACCTGCTGATTTTCATTCGCCATCAACCTCA	60
QY	261	GATTGAGACCAATGTGATCCACTCTGCATATGAGCAGGTGTGAAGAAGCTTCTCTTCT	320
DB	61	AATCAGACCAATGTCTCATCGATTCGATATGCAATGCGAATGCTGAAGAACTGTGTTTT	120
QY	321	TGGATCATCTGCAATTTACCTTAAATTTGCTCTCAGCAATCTCTGAGTCTGCTTTGTT	380
DB	121	GGGTTCTCTTGGCAATTTACCCCAATTTGACCCCAACCGATTCGGAAGATGCTTTGCT	180

QY 381 AACAGCATCGCTTGAACCAACTAATGAGTGGTATGCTATGCTAAGATCGCTGGGATTA 440
 Db 181 TACTGACCCCTTAGAGCCACAAATGAATGGTATGCCATTGCCAAGATTGCTGGGATCA 240
 QY 441 GACTTGTCCAGGCTTATAGGATTCAGACCGATGGGATGCAATCTCTGGCATCGCTACTAA 500
 Db 241 AATGTGCCAGGCTTACAGAAATTCAGATAAGTGGGATGCAATTCGGGAATGCCACCAA 300
 QY 501 TCTCTATGTCTTAATGACAATTTCCACCGGAGAATTTCTCATGTCTTCTCTGCTCTTAT 560
 Db 301 CTTATATGACCATATGACAAATTTTCATCCGAGAAATTCGATGTGTACCTGCTCTAT 360
 QY 561 GAGGAGGTTCCAGAGCGGAAAGTGAATGGAGCGGAGAGTTGTGTGTGGGTACAGG 620
 Db 361 GAGAAGGTTTCATGAGCGCAAGGTCAATGGTGCTAAGGAGGTGGTGTGTGGGACCGG 420
 QY 621 TAGTCCGTTGAGGAGTCTTTCATGCTTTCATGATTTGGCTGATGCTTGTGTTTCTTGT 680
 Db 421 AAGTCCATTTGAGGAGTCTTTCATGCTTTCATGATTTGGCAGATTTGGCAGCGGTTGTCTTATGAT 480
 QY 681 GGATCATACAGCGGTTGGAGCATGTTAACTTGAAGTGGTCAAGAGTGAATATTAG 740
 Db 481 GGAAGATATAGCGGACTGGAGCATTTGAATGTAGGAGTGAAGAGGTTACTATTAA 540
 QY 741 AGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGAAAGCTTGGATGGGATG 800
 Db 541 GGAATTTGGCTGAGTTGATGAAGGAGTGGTGGGATTTGAAGGGGATCTTGTGTTGGGATTC 600
 QY 801 CACTAAGCCAGATGGCAC 818
 Db 601 TACTAAGCCTGATGGGAC 618

RESULT 8
 AW041037
 LOCUS EST283901 tomato mixed elicitor, BT1 Lycopersicon esculentum cDNA
 DEFINITION clone cLET8A5, mRNA sequence.

ACCESSION AW041037
 VERSION AW041037.1 GI:5899791
 KEYWORDS EST.
 SOURCE tomato.

ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 1. (bases 1 to 652)
 D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
 Roming,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,
 Fraser,C.M., Venter,J.C., Martin,G.B., Tankale,S.D. and Giovannoni,
 J.

REFERENCE Generation of ESTs from tomato leaf tissue
 Unpublished (1999)

TITLE Contact: CUGI
 JOURNAL Clemson University Genomics Institute
 COMMENT 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES Location/Qualifiers
 source
 1..652
 /organism="Lycopersicon esculentum"
 /cultivar="Rio Grande PtoR"
 /db_xref="taxon:4081"
 /clone="cLET8A5"
 /clone_lib="tomato mixed elicitor, BT1"
 /tissue_type="leaf"
 /dev_stage="4-6 week old plants"
 /lab_host="XLI-Blue MRF,"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; cLET - Inoculated with a variety of disease response
 elicitors. Plants exposed to 2,6 dichloroisonicotinic

acid, BTH, jasmonic acid, ethylene, fenthion, BIX,
 okadaic acid, or systemin prior to tissue harvest. EcoRI
 site was destroyed during cloning."
 BASE COUNT 179 a 158 c 118 g 197 t

Query Match 37.9%; Score 354.8; DB 10; Length 652;
 Best Local Similarity 71.8%; Pred. No. 2.1e-93;
 Matches 464; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 7 GACAAATCTGCCAAATCTTCGTGCGGGTCAATCGTGGTTTGGTTGGATTCGCCATTGTC 66
 Db 7 GATATCTCCGTAAGATCTTCGTGCGGGCCACCGTGGACTCGTGGATCGCGCGTGGTT 66
 QY 67 CGCAAGCTTCAGGAACAAGGTTTCAACCAATCTCGTTCTTTAAACACACGCGAGCTTGAT 126
 Db 67 CGGAACCTTTACCAATTTAGGCTGCACAATCTCATCTCCGTACACATTCGATCTCAT 126
 QY 127 CTCATCTCTCAAGCGGATGTTGAATCTCTTTCTTCAAGAGAAGCCAGTTTATGTAATC 186
 Db 127 CTCATTAATCAATCGCGGTGCAATCTCTTCGCGACGAGAAACCTCAATACGTCATC 186
 QY 187 CTAGCAGCAGCTAAAGTTGGTGTATTACGCTTAACACACCTATCTCTGCTGATTTCAAT 246
 Db 187 CTCGACGCGGAAAGTCGCGGCATACACGCAACAATATCTTACCCAGCTGATTTCAATC 246
 QY 247 GGTGTCATCTCCAGATTCAGACCAATGTGATCTGCATATGACACGCGTCTGAAG 306
 Db 247 ACTATAATCTGCAATCCAAACGACGATCGTTTCATCTCTCAATACACAAGTTCAAG 306
 QY 307 AAGCTTCTCTCTTGGATCATCTCTGCAATTTACCTAAATTTGCTCTTCAAGCAATTCCT 366
 Db 307 AAGCTTCTCTCTTGGTTCTTCTATGATTTTACCTAAATTTGCTCTTCAAGCAATTCCT 366
 QY 367 GAGTCTGCTTTGTTAAACAGCATCGTTGAACCAACTAATGATGGTATGCTATTCCTAAG 426
 Db 367 GAAATGCACCTTTTAACTGCTCTTTGGAACCTCAAAATGAATGGTATGCAATTCGCAAA 426
 QY 427 ATCGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACCGATGGATGCAATCTCT 486
 Db 427 ATTGCTGGTATTAATGTGTCAAGCTTATAGATTGAGCATAACTTTGATGCAATTTCA 486
 QY 487 GGCATGCTACTAATCTCTATGTCCTTAATGACAATTTCCACCGGAGAATTCATCATGTG 546
 Db 487 GCAATGCTCAGAAATTTATATGTCAGAAATGACAATTTCCATCTCTGAGAATTCATGTT 546
 QY 547 CTTCTGCTCTTATGAGGAGGTTCCAGAGCGGAAAGTGAATGAGCGCGGAGAAAGTTG 606
 Db 547 TTGCTGCTTTGTTTACGTAGATTTTCATGAAGCAAAAGTTAAACAATCTTTGATAAAGGTT 606
 QY 607 GTGTGGGCTACAGGTAGTTCCTGAGGAGGTTCTTGCATGTTGATG 652
 Db 607 GTGTGGGCTACTGTTCTTCTTTAAGGGAATTTTACATGTCGATG 652

RESULT 9
 BH173899
 LOCUS B216/R RFLP clone from soybean genomic DNA
 DEFINITION clone pB216 R primer, DNA sequence.
 ACCESSION BH173899
 VERSION BH173899.1 GI:15989266
 KEYWORDS GSS.
 SOURCE soybean.
 ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1. (bases 1 to 731)
 Grant,D. and Shoemaker,R.C.
 REFERENCE Unpublished sequence of soybean RFLP probe
 JOURNAL Unpublished (2001)

421 CCGGAAGTCCATTGAGGAGTTCTTGACGTTGACGATTTGGCAGACGCGGTGTCTTCA 480
 677 TGTGTCATCATACAGCGGTTGGAGCATGTTAACATTGGAAGTGTCAAGAAGTACTA 736
 481 TGATGGAAAAGTATAGCGGATTCGAGCATTTGAATGTAGGAGTGGAAAAGGAGTTACTA 540
 737 TTAGAGAGTTGGCTGAGTTGGTGAA 761
 541 TTAAGGAATTGGCTGAGTTGATGAA 565

RESULT 11
 BQ116267 752 bp mRNA linear EST 22-JUL-2002
 LOCUS EST601830 mixed potato tissues Solanum tuberosum cDNA clone STMDG80
 DEFINITION 5' end, mRNA sequence.

ACCESSION BQ116267
 VERSION BQ116267.2 GI:21917394
 KEYWORDS EST.

SOURCE potato.

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 752)

AUTHORS Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,

Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and

Karamycheva, S.A.

Generation of a set of potato cDNA clones for microarray analyses
 Unpublished (2002)
 On Apr 17, 2002 this sequence version replaced gi:20168216.

CONTACT: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato@cigr.org

This clone is available through the Research Genetics, contact the
 Research Genetics for further information 1-800-711-6195 or
 cdna@resgen.com

Seq primer: T3.

FEATURES Location/Qualifiers

source

1. .752
 /organism="Solanum tuberosum"
 /cultivar="Kennebec or Binjte"
 /db_xref="taxon:4113"
 /clone="STMDG80"
 /clone_lib="mixed potato tissues"
 /tissue_type="mixed tissues"
 /lab_host="SOLR"

/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; supplier: Combination of untreated and Phytocphora
 infestans-treated libraries of stolons, leaves, leaflets,
 axillary buds of stem explants, petioles, germinating eyes
 of tubers, or roots."

BASE COUNT 207 a 184 c 140 g 221 t

ORIGIN

Query Match 37.5%; Score 350.6; DB 14; Length 752;
 Best Local Similarity 69.9%; Pred. No. 4e-92;
 Matches 501; Conservative 0; Mismatches 214; Indels 2; Gaps 2;

7 GACAAATCTGCCAAAATCTTCGTCGGGGTTCATCGTGGTTGGTTGGATCTGCCATTGTC 66
 35 GATACTTCGCTAAGATCTTCGTCGGGTACCGTGGACTCGTCGGATCCGCGGTGTT 94

67 CGGAACCTTCAGGAACAAGTTTACCAATCTCGTTCTTAAACACACCGCGAGTTGAT 126
 95 CGGAACCTTTACCAATTAGGTGTCACAAATCTCTCCGTCACACATTCGGATCTCGAT 154

127 CTCACCTCGTCAAGCCGATGTGAATCTCTTTCTTCAAGAGAGCCAGTTTATGATC 186
 155 CTCACCTAACCAATCCCGCTCGAATCTTCTTCGCGGACGAGAAACCTCAATACGTCATC 214

187 CTAGCAGCAGCTAAAGTTGGTGGTATTTCACGTAACACACCTATCTCTGCTGATTTCATT 246
 215 CTCGCGCGCGAGGTTCGGCGCATACAGCAACAATACTTACCAGCTGATTTCATC 274
 247 GGTGTCAATCTCCAGATTTCAGACCAATGTGATCCACTCTGCAATATGAGACCGGTGGAAG 306
 275 ACTATAAATCTTCAAAATCCAAACAAACGTCATCGTTTCATCTTCAATCACAAAGTTCA 334
 307 AGCTTCTCTCTCTTGGATCATCTTCCTGCAATTTTACCCCTAAATTTGCTCCCTCAGCAATTCCT 366
 335 AAGCTTCTCTCTTGGTTCCTTATTTTACCCCTAAATTTGCTCCCTCAGCAATTTCCC 394
 367 GAGTCTGCTTTTAAACAGCATCGCTTGAAACCAACTAATGAGTGTGATGCTATTGCTAAG 426
 395 GAAATGACATTTTAACTGCCCTTTGGAACCTACAATGATGTATGCAATTCGAAA 454
 427 ATCGCTGGGATTAAGACTTTGTCAGGCTTATAGGATTCAGACGGATGGGATCAATCTCT 486
 455 ATTGCTGGTATCAAAATGTGTCAAGCTTATAGATTGCAGCATAACTTTGATGCAATTTCA 514
 487 GGCATGCGCTACTAATCTCTATGGTCTCTAATGACAAATTTCCACCCGGAGAAATTTCTCATGTG 546
 515 GCAATGCCACGAAATTTATAGTACGAATGACCAATTTCCATCTCTGAGAAATTTCTCATGTT 574
 547 CTTCCTGCTCTTATGAGGAGTTTCCACGAGCGGCAAGTGAATGGAGCGGAGGAGTTGTG 606
 575 TTGCTCTGCTTTGTTACGTAGATTTTCATGAAGCAAAAGTAACTCTTGATAAAGTTGGTT 634
 607 GTGTGGGG-TACAGTAGTCCGTTGAGGAGTTCTTGGCATGTTGATGATTTGGCTGATGC 665
 635 GTGTGGGAAACTGCTTCTCTTTAAGGGAATTTTACATGTCGATGATTTAACCGATGC 694
 666 TTGCTGTTTCTGCTGGATCGATCAGCGGGTTGGAGCATGTTACATTTGGAAGTGG 722
 695 AGTAGTGTTTTGGTGGAGAAATTATAGTGAATTAGA-CATGTTTATGTTGGGAGTGG 750

RESULT 12

LOCUS BQ440809

DEFINITION GA_Ea0010D16f Gossypium arboreum 7-10 dpa fiber library Gossypium

ACCESSION BQ440809

VERSION BQ440809.1 GI:13350461

KEYWORDS EST.

SOURCE Gossypium arboreum.

ORGANISM Gossypium arboreum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.

REFERENCE 1 (bases 1 to 687)

AUTHORS Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry

, D., Wood, T.C., Leslie, A. and Wilkins, T.A.

An integrated analysis of the genetics, development, and evolution

of the cotton fiber

Unpublished (2000)

JOURNAL

COMMENT

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAATACGACTCACTATAGGG

High quality sequence stop: 683.

Location/Qualifiers

1. .687

/organism="Gossypium arboreum"

/strain="AKA"

/cultivar="8400"

/db_xref="taxon:29729"

/clone="GA_Ea0010D16f"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"


```
Db 605 GGTACTGGT 613
|||||
A1726948 664 bp mRNA linear EST 11-JUN-1999
LOCUS BNLGH16903 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
DEFINITION to (AF045286) GDP-4-keto-6-deoxy-D-mannose-3
5-epimerase-4-reductase [Arabidopsis thaliana], mRNA sequence.
ACCESSION A1726948
VERSION A1726948.1 GI:5045800
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 664)
AUTHORS Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
TITLE ESTs from developing cotton fiber
JOURNAL Unpublished (1999)
COMMENT Contact: Ben Burr
Biology Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burrb@bnl.bnl.gov
Seq primer: T3 Primer.
Location/Qualifiers
1..664
/organism="Gossypium hirsutum"
/cultivar="Acala Maxxa"
/db_xref="taxon:3635"
/clone_lib="Six-day Cotton fiber"
/tissue_type="immature fiber"
/dev_stage="Six days post anthesis"
/lab_host="XLI-Blue"
/notes="Vector: pBluescript II KS-"; 2 others
BASE COUNT 161 a 204 c 128 g 169 t
ORIGIN
Query Match 34.8%; Score 326; DB 9; Length 664;
Best Local Similarity 71.7%; Pred. No. 6.6e-85;
Matches 439; Conservative 0; Mismatches 172; Indels 1; Gaps 1;
QY 2 TGTTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTATCGGCA 61
Db 54 TCTCGAACAAATCAGCGAAGATCTTCGTANCAAGCCATCGAGCCCTAGTTGTTCTGCCA 113
QY 62 TTGTCGCAAGCTTCAGGAACAGGTTTCCACCAATCTCGTTCTTAAACACACGCGGAGC 121
Db 114 TANTTCGTAACTTCAGTCCCTAGGCTTTACCAACCTCTCTACTCCGACCCATCGCGATC 173
QY 122 TTGATCTCACTCGTCAAGCGGATGTTGAATCTCTTTCTCAAGAGAGCCAGTTTATG 181
Db 174 TGGACCTTACTCGCAATCGAGTCGATCTCTTTCGCCGATGAGAACCTCACTAGC 233
QY 182 TTAATCTCAGCAGCTAAAGTTGGTGTATTACGCTAACACACTATTCCTGCTGAT 241
Db 234 TCGTACTAGTCGCGCTAAAGTTGGTGGATCCACGCCAACACACTTACCCCTGCGGAT 293
QY 242 TCATTGGTGTCAATCCAGATTTCAGACCAATGTGATCCACTCTGCATATGAGCGG 301
Db 294 TCATTGCCATCACTCCAAATCCAGACCAAGTCATGATTCCTCTTACCGCCAGCGG 353
QY 302 TGAAGAAGCTTCTCTCTTGGATCATCTGCAATTTACCTAAATTTGCTCTCAGCCAA 361
Db 354 TTAACAAATCTCTCTCTCGGTTCTCTTGATTTATCCCAAGTTTGGCGCGCAACCA 413
QY 362 TTCTGAGTCTGTTGTTTAAACAGCATCGCTTGTGAACCAACTAATGATGATGATTTG 421
|||||
Db 414 TCCCTGAAATCGCTCTTTATCCGGTCCCTCGAACCCACCAACGAATGATGCGGTTG 473
QY 422 CTAAGATCGCTGGATTAAAGACTTGTACGGCTTATAGGATTCAGACCGATGGATGCAA 481
Db 474 CCAAGATCGGGGATCAAAATGTCCAGCGGTACAGAAATTCAGCACGGCTTCGATGCTA 533
QY 482 TCTCTGGCATGCTACTATCTATCTATGCTCTTAATGACAAATTTCCACCCGGAGAAATCTC 541
Db 534 TTTCCGCTATGCGGCAAAATTTGTACGGCCCCCAACGATTAATTTCCATCTCTGAAATTTCC 593
QY 542 ATGTGCTTCTCTCTCTTATGAGGAGTTCCACAGCGGAAAGTGAATGAGCGGAGGAAG 601
Db 594 ACGTTTTCGCCGCTTTGATGCGGAGGTTCC-TAAAGCCAAAGTGGATAGACGCAAAAAAAG 652
QY 602 TTGTGGTGGG 613
Db 653 TAGTTGTGGG 664
RESULT 15
A1487671 590 bp mRNA linear EST 18-MAY-2001
LOCUS EST245993 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
DEFINITION CLED13M21, mRNA sequence.
ACCESSION A1487671
VERSION A1487671.1 GI:4383042
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 590)
AUTHORS Alcalá,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato carpel tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
Location/Qualifiers
1..590
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLED13M21"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-Blue MRF"
/notes="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
XhoI; CLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 166 a 151 c 96 g 177 t
ORIGIN
Query Match 34.4%; Score 322.2; DB 9; Length 590;
Best Local Similarity 72.5%; Pred. No. 8.1e-84;
Matches 417; Conservative 0; Mismatches 158; Indels 0; Gaps 0;
QY 13 TCTGCCAAATCTTCGTCGGGTCATCGTGGTTGGTATCGCATCTGCGCAAG 72
Db 5 TCCGCTAAGATCTTCGTCGGGCCACCGTGGATCTCGTCGATCGCGGTTGTTGGA 64
QY 73 CTTTCAGGAACAGGTTTCCCAATCTCTTTTAAACACACGCGAGCTGATCTACT 132
|||||
```

```
Db      65  CTTTACCAATTAGGTCGACAAATCTCATCTCCGTACACATTCGGATCTCGATCTCACT 124
QY      133  CGTCAAGCCGATGTTGAATCCTCTTTCTCAAGAGAGCCAGTTTATGTAATCCTAGCA 192
Db      125  AATCAATCCGCGTCAATCTCTTCGCGGACGAGAAACCTCAATACGTCATCCTCGCA 184
QY      193  GCAGCTAAAGTTGGTGTATTACGCTACACACCTATCCTGCTGATTTCAATGGTGTG 252
Db      185  GCGCGGAAAGTCGGCGGCATACACGCAACAATACTTACCCAGCTGATTTCACTATA 244
QY      253  AATCTCCAGATTACAGACCAATGTGATCCACTCTGCAATATGAGCAGGTTGTGAAGAGCTT 312
Db      245  AATCTGCAATCCAAACGAACTTATCGTTTCATCCTTCAATCACAAGTTCAGAGCTT 304
QY      313  CTCTTCCTTTGGATCATCTCTGCAATTTACCTAAATTTGCTCCTCAGCCAAATCCTGAGTCT 372
Db      305  CTGTTCTCTTGGTCTTTCATGTATTTACCTTAAATTTGCTCCTCAACCAATTCCTGAAAT 364
QY      373  GCTTTGTTTAAACGATCGCTTGAACCACTAATGATGGTGTGATGCTATGCTAAGATCGCT 432
Db      365  GCACCTTTAACTGCTCTTTTGAACCTTCAAAATGAATGGTATGCAATTCGAAATTTGCT 424
QY      433  GGGATTAAAGCTTGTGAGGCTTATAGATTTCAGACGGATGGATGCAATCTCTGSCATG 492
Db      425  GGTATTAAATGTGTCAAGCTTATAGATTGCAAGCATAACTTTGATGCAATTTTCAAGCATG 484
QY      493  CCTACTAATCTCTATGGTCTTAATGACAAATTTCCACCCGAGAAATTCATGTGCTTCT 552
Db      485  CCTACGAATTTATATGGTACGAATGACAATTTCCATCCTGAGAAATTCATGTGCTTCT 544
QY      553  GCTCTTATGAGGAGGTTCCACGAGGCGAAAGTAA 587
Db      545  GCTTTGTTACGTAAATTTTCATGAACAAAAGTTAA 579
```

Search completed: June 3, 2003, 04:00:13
Job time : 1678 secs

QY 306 GAAGCTTCTCTTCTGGATCATCTGCAATTTACCTAAATTTGCTCTCAGCAATTC 365
Db 1704549 GCGGCTCTCTTCTGGGCTCGTGCATCTACCCGAACTCGCCGCGAGCGATCCC 1704608
QY 366 TGAGTCTGCTTTGTTAAACAGCATCGCTTGAACCACTAATGAGTGTATGCTATTGCTAA 425
Db 1704609 GGAGAGCGGCTGCTCAACCGGTCGTTGGAGCGGCAACAGAGCGGTACGCGATCGCCAA 1704668
QY 426 GATCGCTGGGATTAAGACTTTGTGAGGCTTATAGATTCAGCAGCGATGGGATGCAATCTC 485
Db 1704669 AATCGCGGATTTCTGGGTCAGGCGTCCAGGCGTGCAGCGCAATGGCTGCGGTGATCTC 1704728
QY 486 TGGGATGCTTAACTCTATGCTCTAATGCAATTTCCACCGGAGAAATTTCTCATGT 545
Db 1704729 GGGGATGCTTAACTCTATGCTCTAATGCAATTTCCACCGGAGAAATTTCTCATGT 545
QY 546 GCTTCTGCTCTTATGAGGAGTTTCCAGGAGGCAAGTGAATGGAGCGGAGAAATTTGT 605
Db 1704789 GCTGCGGCACTCATCGCGCTATGACGAGGCAAGCCAGTGGCGCGCCCAACGTGAC 1704848
QY 606 GGTGTTGGGTACAGGTAGTCCGTTGAGGAGGATTTCTGCAATTTGATGATTTGGCTGATGC 665
Db 1704849 CAATGGGCGACCGGCAACGCGCGGAGTTGCTGCACTGCACTGCACTGCGGAGCGC 1704908
QY 666 TTGTTGTTTCTGCTGATCGATACAGCGGTTGGAGCATGTTAACTTGAAGTGGTCA 725
Db 1704909 ATCCCTGTATCTGTGAACATTTCCAGCGGCGAGCCCATGTCAACGTGGAAACCGCAT 1704968
QY 726 AGAAGTCACTATTAGAGAGTTGCTGATGTTGGAAGAGTTGTTGGTTTGAAGGAA 785
Db 1704969 CGACCAACCATCGGCGAGATCCCGGAGTGGTGGCTCGGCGGTATAGCGGCA 1705028
QY 786 GCTTGGATGGATTTGACATAAGCCAGATGGCAGCAGGAGAACTTATGACAGCTCAAA 845
Db 1705029 AATCCGCTGGATCCAGCAACCGGAGCAACAGCACTGCTGATGTTTGGT 1705088
QY 846 GCTCGGCTTTGGTTGGACACTAAGTTTCTTTAGAGATGGTCTGAGCCAAACTTA 905
Db 1705089 GCTACGGGAGCGGATGGCGCTTCGATCGGCTCGCGACGCGATCGAGCGACGT 1705148
QY 906 TGATTGGTATTTGAAGATG 925
Db 1705149 GCGGTGATTCGCGAGCAG 1705168

RESULT 2

US-09-060-756-1
; Sequence 1, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; FILE INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 12732
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-1

Query Match 29.4%; Score 275.2; DB 4; Length 12732;
Best Local Similarity 56.2%; Pred. No. 1.1e-82;
Matches 517; Conservative 0; Mismatches 403; Indels 0; Gaps 0;
QY 6 TGACAAATCTGCCAAATCTTCTGCGGGTATCGCTGTTGGATCTGCCATTTGT 65

Db 8106 TGACCGCGCGCGCGGCTTACATCGCCGGGATCGCGGCTGGTCCGGTCCGCGTGTCT 8165
QY 66 CGCAAGCTTTACAGAACAAAGTTTACCAATCTCGTTCTTAAACACACCGCGAGCTTGA 125
Db 8166 ACACAGCTTTGCGGCGCGGGTTTACCAACTCTGCTGGTTCGCGTACGCGCGAGCTTGA 8225
QY 126 TCTCACTCTCAAGCCGATGTTGAATCTCTTTCTCAAGAGAAAGCAGTTTATGTAAT 185
Db 8226 TCTGACGGATCGCGCGCGGACGTTGACITCTGTTCTCGAGTCGAGSCCGCAGGTCTCAT 8285
QY 186 CTTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTTAACAACACTATCTCTGCTGATTTTCA 245
Db 8286 CGACGCGCGCGCGCGGTCGCGGATCTCTGCGCAACGACCTACCGCGGATTTCTCT 8345
QY 246 TGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATAGCACGCGTGA 305
Db 8346 GTCGGAAACCTCCAGATCCAGTCAACCTGTGATCGCGCGGTCGCGCGGTCGCC 8405
QY 306 GAAGCTTCTCTTCTTGGATCATCTGCAATTTACCTAAATTTGCTCTCAGCAATTC 365
Db 8406 GCGGCTGCTGTTCTTGGGCTCGTGTGATCTACCGAACTCGCCCGCGAGCGATCCC 8465
QY 366 TGAGTCTGCTTTGTTAAACAGCATCGCTTGAACCAACTAATGAGTGGTATGCTATTGCTAA 425
Db 8466 GAGAGCGCGCTGCTCACCGTCCGTTGGAGCGGCAACGACGCGGTACGCGATCGCCAA 8525
QY 426 GATCGCTGGGATTAAGACTTTGTCAGGCTTATAGGATTCAGCAGCGATGGATGCAATCTC 485
Db 8526 AATCGCGGATCTCTTGGGTCAGCGGTCGCGCGCAACATGGCTCCGCTGGATCTC 8585
QY 486 TGCGATGCTCTAATCTCTATGCTCTAATGACAAATTTCCACCGGAGAAATTTCTCATGT 545
Db 8586 GCGGATGCTTCTGCTGCACTGTCAGGCGGAGGCAACTTTTCGCGTCCGGCTCGCATCT 8645
QY 546 GCTTCTGCTCTTATGAGAGGTTTCCACGAGCGGAAAGTGAATGAGCGGAGAGATTTGT 605
Db 8646 GCTGCGGCACTCATCCGCGCTATGACGAGGCGCAAGCCAGTGGCGCGCCCAACGTGAC 8705
QY 606 GGTGTTGGGTACAGGTAGTTCGTTGAGGAGTTTCTTGCATGTTGATGATTTGCTGATGC 665
Db 8706 CAATCGGCGCACCGGACCGCCCGCGAGGATTTGTCACGTCGACGACTGGCGAGCGC 8765
QY 666 TTGTTGTTTCTTGTGATCGATACAGCGGTTGGAGCATGTTAACTTGAAGTGGTCA 725
Db 8766 ATGCTGCTATCTGCTGGAACATTTTCAGCGGCGGACCCCATGTCACGTTGGAACCGCAT 8825
QY 726 AGAAGTGAATTAAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGGAA 785
Db 8826 CGACCAACCATCGCGGAGATCGCGGATGTTGCTCGCGGTAGGCTATAGCGGCGA 8885
QY 786 GCTTGGATGGGATTCACATTAAGCCAGATGGCACACCGAGGAACTTATGACAGCTCAAA 845
Db 8886 AATCCGCTGGGATCCAGCAACCGGAGGAAACACGACCAACTGCTGGATGTTTGGT 8945
QY 846 GCTCGGCTTTGGTTGGACACCTAAGGTTTCTCTTGAAGATGGTCTGAGCCAAACTTA 905
Db 8946 GCTACGGGAGCGGATGGCGGCTTCGATCGGCTGCGGAGCGCATCGAGCGGACGCT 9005
QY 906 TGATTGGTATTTGAAGATG 925
Db 9006 GCGGTGATTCGCGAGCAG 9025

RESULT 3

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.

;; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

;; FILE OF INVENTION: TUBERCULOSIS

;; FILE REFERENCE: 24366-20007.00

;; CURRENT APPLICATION NUMBER: US/09/103,840A

;; CURRENT FILING DATE: 1998-06-24

;; NUMBER OF SEQ ID NOS: 2

;; SOFTWARE: PatentIn Ver. 2.1

;; SEQ ID NO 1

;; LENGTH: 4411529

;; TYPE: DNA

;; ORGANISM: Mycobacterium tuberculosis

;; OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match 29.4%; Score 275.2; DB 4; Length 4411529;
Best Local Similarity 56.2%; Pred. No. 5e-81;
Matches 517; Conservative 0; Mismatches 403; Indels 0; Gaps 0;

QY 6 TCACAAATCTGCAAAATCTTCGTCCGGGTATCATCGTGGTTGGTTGGATCTGCCATTGT 65
DB 1704120 TCACCGCGCGCCGGGTCTACATCGCGGGATCGCGCCCTGGTGGGTCCGCGTCTGT 1704179

QY 66 CGCAGCTTTCAGGAACAGGTTTCACCAATCTCTGTTCTTAAACACACACGCGAGCTTGA 125
DB 1704180 ACGACGTTTGGCGCGGGGTTTACCAACTCTGTGTCGGTTCACGCGCCGAGCTTGA 1704239

QY 126 TCTCACTCGTCAAGCCGATGTTGAATCTCTTCTTTCTCAAGAGAAGCCAGTTTATGTAAT 185
DB 1704240 TCTGACGGATCGGCGCGACGTTCTGACTTCTTCGAGTCCGAGCGCAGTCTGTAT 1704299

QY 186 CTTAGCAGCAGTAAAGTTGGTGTATTCACGCTAACACACTATCTCTGCTGATTTCAT 245
DB 1704300 CGACCGCGCGCCGGGTCCGGCGGATCTCTGTCACACGACACTACCGCGGATTTCT 1704359

QY 246 TGGTGTCAATCTCCAGATTCAGACAAATGTGATCACTCTGCATATAGCACGGTGTGAA 305
DB 1704360 GTCCGAACCTCCAGATCCAGTCACTGTGTGATGCCGCGCGGTGGTCC 1704419

QY 306 GAAGCTTCTCTTCTTGGATATCTCTGATATTCCTTAAATTTGCTCTCAGCCCAATTCC 365
DB 1704420 CGGCTGTCTTCTCTGGGCTGTCTGTGATCTACCGAACTCGCCCGCAGCCGATCC 1704479

QY 366 TGAGTCTGTTTGTAAACAGCATCGTTGAACCACTAATGATGTTGTTATGCTTAA 425
DB 1704480 GGAGAGCGCGTCTCACCGTCCGTTGGAGCGCACCAACGCGGTACGCGATCGCCAA 1704539

QY 426 GATCGTGGGATTAAGACTTGTGAGCTTATAGGATTCAGACGGATGGATGCAATCTC 485
DB 1704540 AATCCCGGCATCTTGGCGGTCCAGCGGTGCGCCCAACATGGCTCCGCTGGATCTC 1704599

QY 486 TGGCATGCCCTACTAATCTCTATGTCCTAATGACAAATTTCCACCGGAGAAATCTCATGT 545
DB 1704600 GCGATGCCCAACCTGTACGGCGAGCGCAACTTTTCGCGTCCGCTCGCATCT 1704659

QY 546 GTTCTGTCTTATGAGAGGTTTCCAGAGCGGAAAGTGAATGAGCGGAGGAATGTGT 605
DB 1704660 GCTGCCGCACTCATCCGCGCTATGACGAGGCGCAAGCAGTGGCGCGCCCAAGCTGAC 1704719

QY 606 GGTGTGGGTACAGGTAGTCCGTTGAGGAGTCTTGCATGTTGATGTTGGCTGATGC 665
DB 1704720 CAACTGGGGCACCGGACCGCCCGCAGCGGAGTTGTGTCACGTCGACGACTCGCGAGCGC 1704779

QY 666 TTGTGTTTTCTTGCTGATTCGATACAGCGGGTTGGAGCATGTTAAACATTGGAAGTGGTCA 725
DB 1704780 ATGCTGTATCTGCTGGAACATTTCCAGCGGCGGACCCCATGTCAACGTGGGAACCGCAT 1704839

QY 726 AGAAGTACTATAGAGATTCGCTGAGTTGTGTGAAGAGTTGTTGTTTTGAAGGAA 785
DB 1704840 CGACACACCATCGGCGAGATCGCGGAGATGTTGCTCGCGGTAGGTTATAGCGGCA 1704899

QY 786 GCTTCGATGGGATTCGCACTAAGCCAGATGGCACACCGAGGAACCTTATGACAGCTCAA 845
DB 1704900 AACCCGCTGGGATCCAGCAAAACCGGACCGGAACACACGACAACTGCTGGATGTTTCGGT 1704959

QY 846 GCTCCGCTTTTGGTTGGACACCTAAAGTTTCTTTAGATAGTGTCTGAGCAAACTTA 905
DB 1704960 GCTACGGAGGCGGATGCGGCCCTTCGATCGCTGCTGCGGACGCGATCGAGCGACGCT 1705019

QY 906 TCATTGGTATTGAAGAATG 925
DB 1705020 GCGGTGGTATCGGAGCAG 1705039

RESULT 4

US-09-453-702B-121/c
; Sequence 121, Application US/09453702B
; Patent No. 6365723

GENERAL INFORMATION:

;; APPLICANT: Blattner, Frederick R.
;; Burland, Valerie
;; Perna, Nicole T.
;; Plunkett, Guy
;; Welch, Rod
;; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;; NUMBER OF SEQUENCES: 265
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Quarles & Brady
;; STREET: 1 South Pinckney Street
;; CITY: Madison
;; STATE: WI
;; COUNTRY: US
;; ZIP: 53701-2113

COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Word Perfect 8.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/453,702B
;; FILING DATE: 03-Dec-1999
;; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 60/110,955
;; FILING DATE: 04-DEC-1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seay, Nicholas J.
;; REGISTRATION NUMBER: 27386
;; REFERENCE/DOCKET NUMBER: 960296.95017
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (608) 251-5000
;; TELEFAX: (608) 251-9166
;; INFORMATION FOR SEQ ID NO: 121:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 14187
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)
;; SEQUENCE DESCRIPTION: SEQ ID NO: 121:

US-09-453-702B-121

Query Match 28.8%; Score 269.8; DB 4; Length 14187;
Best Local Similarity 57.1%; Pred. No. 8.3e-81;
Matches 530; Conservative 0; Mismatches 372; Indels 27; Gaps 1;

QY 22 ATCTCGTCGCGGGTCATCGTGGTTGGTTGGATCTGCCATTGTCCGCAAGCTTCAGGAA 81
DB 6374 ATTTTATTTGTTGGTCCACCAAGGATGGTTGGATCAGTTATACCGAGCGCTCAACAA 6315

QY 82 CAAAGTTTCAACAATCTCGTCTTAAACACACGCGAGCTTGATCTCTACTCGTCAAGCC 141
DB 6314 CGTGATGATGTTGAGTTGGTTTACGTACTCGGGATGAATTGAACTTGTGGATAGTAGC 6255

QY 142 GATGTTGAATCTCTCTTTCTCAAGAGAGCCAGTTTATGTAATCTTAGCAGCAGTAA 201
DB 6254 GCTGTTTGGATTTTTTTTTTCTTTCACAGAAATTCGACCAGGTTTATTTTGGCAGCAGCAAA 6195

APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: ANTIGENS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,972
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PP-0400 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1352 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BLADNOT04
CLONE: 1318190
US-08-937-972-4

Query Match 11.0%; Score 103.2; DB 2; Length 1352;
Best Local Similarity 48.9%; Pred. No. 8.4e-25;
Matches 341; Conservative 0; Mismatches 348; Indels 9; Gaps 2;

QY 124 GATCTCACTCGTCAAGCCGATGTTGAATCCTTCTTCTCAAGAGAGCCAGTTTATGTA 183
DB 223 GATCTCAGGATACAGACAGCCGCCGCTGTTTGAGAGGTCCAACCCACACAGTC 282
QY 184 ATCTAGACGACGCTAAAGTTGGTGTATTACGCTTAACACACCTATCTCTGCTGATTC 243
DB 283 ATCCATCTTGTGCAATGTTGGGGGCGCTGTTCCGGAATATCAATATCAATTTGGACTTC 342
QY 244 ATTGGTGTCAATCTCAGATTAGACCAATGATGATCCATCTGTCATATGACGCGGTG 303
DB 343 TGAGAGAAAAGTGACATGAACGACAGCTGCTGCACTCGGCTTTGAGGTGGCGCC 402
QY 304 AAGAAGCTTCTCTTCTTGGATCATCTGCAATTAACCTAAATTTGCTCTCAGCAAT 363
DB 403 CGCAAGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 462
QY 364 CTGAGTCTGCTTGTGTTAAGACGATCGCTGTAACCAACTAATGAGTGTATGCTATTGCT 423
DB 463 GATGAGACCATGATCCAAATGGGCTCCCAACACAGCAATTTGGGTACTCGTATGCC 522
QY 424 AGATCGCTGGGATTAAGACTTGTGAGCTTATAGGATTCAGCAGGATGGGATGCAATC 483
DB 523 AAGAGGATGATGACGTGAGAAACAGGGCTACTTCCAGCAGTACGGCTGCACCTTCACC 582
QY 484 TCTGGCATCGCTACTTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
DB 583 GCTGTCTATCCCCCAACAGCTCTCGGGGCCCCAGCAACTTCAACATCGAGATGCCAC 642

544 GTGCTTCTGCTCTTATGAGGAGGTTCCACGAGGCGAAAGTGAATGAGCGGAGGAAGTT 603
DB 643 GTGCTGCTTGGCTCTATCCACAAAGGTGCACCTGGCGCAAGAGCAGCGGCTCGGCC 699
QY 604 GTGGTGTGGGTACAGGTAGTCCGTTGAGGAGGTTCTTTCATGATGATTTGGCTGAT 663
DB 700 ACGGTGTGGGTACAGGAATCCGCGGAGGAGTTCATATATCTCGTGGACCTGGCCCC 759
QY 664 GCTTGTGTTTTCTTGTGATCGATACACGCGGTTGGAGCATGTTA-----ACATTGGA 717
DB 760 CTCCTTATCTGCTGCTGCGGAGTACAATGAAGTGGAGCCCATCATCTCTCGGTGGC 819
QY 718 AGTGTCTAAGAGTCACTATTAGAGAGTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTT 777
DB 820 GAGGAAGATGAGGTCTCCATCAAGGAGGCGCGGCGGTGGTGGAGGCCCATGGAATTC 879
QY 778 GAAGGGAAGCTTGGATGGATTGCACTAAGCCAGATGG 815
DB 880 CATGGGAAGTCACTTTGATACAACCAAGTCGGATGG 917.

RESULT 7
US-09-221-017B-669/c
Sequence 669, Application US/09221017B
Patent No. 644799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 669:
SEQUENCE CHARACTERISTICS:
LENGTH: 517 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)

```

; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...517
US-09-221-017B-669

Query Match          5.8%; Score 54.6; DB 4; Length 517;
Best Local Similarity 52.9%; Pred. No. 1.5e-08;
Matches 117; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 703 CATGTTAAACATTGGAAGTGGTCAAGAGTCACTATTAGAGAGTTGGCTGAGTTGGTGA 762
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 CATATCAACATTGGCACC GGAGGAGAAATCTCAATTCGGATTGGCTTCGCTGATGCA 366

QY 763 GAGGTTGTTGGTTTGAAGGAGCTTTGGATGGATTGCACTAAGCCAGATGCCACCG 822
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 AAGACCATCGGTTACGAGGGCCTTATCGAGTTGCACTCTCCAAACCGGACGCGATG 306

QY 823 AGGAACCTATGACAGCTCAAGCTCGCTCTTTGGTTGGACACCTAAGCTTCTCTT 882
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 CGCAACTGACGATGTAAGCAAGCTCCATGCTTAGGCTGGAAGCATCGGATCGACATA 246

QY 883 AGAGATGGTCTGAGCCAAACTTATGATTGGTATTTGAAGAA 923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 ACCACAGGTGACGAATGTTAGTGATTCGTCCGAA 205
```

RESULT 8

```

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22133-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match          4.4%; Score 41.4; DB 1; Length 7218;
Best Local Similarity 2.3%; Pred. No. 0.0026;
Matches 9; Conservative 214; Mismatches 160; Indels 0; Gaps 0;

QY 552 TGCTTTATGAGGAGTTCCAGGAGCGAAAGTGAATGAGCGGAGGAGTTGCTGTTG 611
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1440 TGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1381

QY 612 GGGTACAGTAGTCGTTGAGGAGTCTTTCATGCTTCATGATTGGCTGCTTGTGT 671
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1380 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1321

QY 672 TTTCTTGCTGATGATACAGCGGTTTGGAGCATGTTTAAACATTGGAAGTGGTCAAGA 731
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1320 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1261

QY 732 GACTATTAGAGATTGGCTGAGTTGGTGAAGAGTTGTTGTTTGAAGGAACCTTGG 791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1260 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1201

QY 792 ATGGATTGCACTAAGCCAGATGGCACACGAGAAACTTATGGACAGCTCAAGCTCGC 851
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1200 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1141

QY 852 GTCTTTGGTTGGACACCTAAGTTTCTTTAGAGATGCTGCTGAGCCAAACTTATGATTG 911
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1140 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081

QY 912 GTATTGGAAGATGTTTCCAACC 934
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1080 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1018
```

RESULT 9

```

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22133-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
```


/ FILING DATE: 16-MAY-1979
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Dianne Casuto
/ REGISTRATION NUMBER: P-40,943
/ REFERENCE/DOCKET NUMBER: 4952.US.P2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (847)-938-3137
/ TELEFAX: (847)-938-2623
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 925 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ US-09-078-166-1

Query Match 3.9%; Score 36.6; DB 3; Length 925;
Best Local Similarity 47.8%; Pred. No. 0.029;
Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 573 CGAGGCGAAAGTGAATGAGCGGAGGAAAGTGTGGTGTGGGTACAGGTAGTCCGTTGAG 632
| | | | |
Db 754 CGAGGCGGTGAGGTTGGTGTGGGCCCGATTTCCAGGTAGTCCGTTGAGT 695

QY 633 GGAGTTCTTTCATGTTGATGATTTGGCTGATGCTTTCTTCTGCTGGATCCGATACAG 692
| | | | |
Db 694 GGAGTGTGAGGTTGGTGTGGGTGTAGCGAGCGGTGCGGTGATGCTGGTCCAGTGGG 635

QY 693 CGAGTTGGAGCATGTTAACTTGAAGTGGTCAAGAAGTGAATTTAGAGAGTTGGCTGA 752
| | | | |
Db 634 TGGTGTGGGTTGGTGGTGTGGCGGTGAGCATGCTGATGAGGGGTGTGTGGGGGT 575

QY 753 GTTGTGAAAGAGTTGTTGGTGTGAAAGGAAAGCTTGGATGGATT 799
| | | | |
Db 574 GGTGGGTGAGGTTGTGAGGGTGGTGTGAAGGGTTGGTGTGATTGAGT 528

RESULT 12
US-08-997-467-1/c
; Sequence 1, Application US/08997467
; Patent No. 6200813
; GENERAL INFORMATION:
; APPLICANT: Katz, Leonard
; APPLICANT: Stassi, Diane L.
; APPLICANT: Summers Jr., Richard G.
; APPLICANT: Ruan, Xiaolan
; APPLICANT: Pereda-Lopez, Ana
; APPLICANT: Kakavas, Stephan J.
; TITLE OF INVENTION: NOVEL POLYKETIDE DERIVATIVES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Rd.
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,467
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: 08/858,003
/ FILING DATE: 16-MAY-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Dianne Casuto
/ REGISTRATION NUMBER: P-40,943
/ REFERENCE/DOCKET NUMBER: 4952.US.P2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (847)-938-3137
/ TELEFAX: (847)-938-2623
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 925 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ US-08-997-467-1

Query Match 3.9%; Score 36.6; DB 4; Length 925;
Best Local Similarity 47.6%; Pred. No. 0.029;
Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 573 CGAGGCGAAAGTGAATGAGCGGAGGAAAGTGTGGTGTGGGTACAGGTAGTCCGTTGAG 632
| | | | |
Db 754 CGAGGCGGTGAGGTTGGTGTGGGCCCGATTTCCAGGTAGTCCGTTGAGT 695

QY 633 GGAGTTCTTTCATGTTGATGATTTGGCTGATGCTTTCTTCTGCTGGATCCGATACAG 692
| | | | |
Db 694 GGAGTGTGAGGTTGGTGTGGGTGTAGCGAGCGGTGCGGTGATGCTGGTCCAGTGGG 635

QY 693 CGAGTTGGAGCATGTTAACTTGAAGTGGTCAAGAAGTGAATTTAGAGAGTTGGCTGA 752
| | | | |
Db 634 TGGTGTGGGTTGGTGGTGTGGCGGTGAGCATGCTGATGAGGGGTGTGTGGGGGT 575

QY 753 GTTGTGAAAGAGTTGTTGGTGTGAAAGGAAAGCTTGGATGGATT 799
| | | | |
Db 574 GGTGGGTGAGGTTGTGAGGGTGGTGTGAAGGGTTGGTGTGATTGAGT 528

RESULT 13
US-08-149-476-281
; Sequence 281, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; EARLIER FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23

[illegible]

1	EARLIER	FILING DATE:	1997-08-22
2	EARLIER	APPLICATION NUMBER:	60/056,662
3	EARLIER	FILING DATE:	1997-08-22
4	EARLIER	APPLICATION NUMBER:	60/056,872
5	EARLIER	FILING DATE:	1997-08-22
6	EARLIER	APPLICATION NUMBER:	60/056,882
7	EARLIER	FILING DATE:	1997-08-22
8	EARLIER	APPLICATION NUMBER:	60/056,637
9	EARLIER	FILING DATE:	1997-08-22
10	EARLIER	APPLICATION NUMBER:	60/056,903
11	EARLIER	FILING DATE:	1997-08-22
12	EARLIER	APPLICATION NUMBER:	60/056,888
13	EARLIER	FILING DATE:	1997-08-22
14	EARLIER	APPLICATION NUMBER:	60/056,879
15	EARLIER	FILING DATE:	1997-08-22
16	EARLIER	APPLICATION NUMBER:	60/056,880
17	EARLIER	FILING DATE:	1997-08-22
18	EARLIER	APPLICATION NUMBER:	60/056,894
19	EARLIER	FILING DATE:	1997-08-22
20	EARLIER	APPLICATION NUMBER:	60/056,911
21	EARLIER	FILING DATE:	1997-08-22
22	EARLIER	APPLICATION NUMBER:	60/056,636
23	EARLIER	FILING DATE:	1997-08-22
24	EARLIER	APPLICATION NUMBER:	60/056,874
25	EARLIER	FILING DATE:	1997-08-22
26	EARLIER	APPLICATION NUMBER:	60/056,910
27	EARLIER	FILING DATE:	1997-08-22
28	EARLIER	APPLICATION NUMBER:	60/056,864
29	EARLIER	FILING DATE:	1997-08-22
30	EARLIER	APPLICATION NUMBER:	60/056,631
31	EARLIER	FILING DATE:	1997-08-22
32	EARLIER	APPLICATION NUMBER:	60/056,845
33	EARLIER	FILING DATE:	1997-08-22
34	EARLIER	APPLICATION NUMBER:	60/056,892
35	EARLIER	FILING DATE:	1997-08-22
36	EARLIER	APPLICATION NUMBER:	60/057,761
37	EARLIER	FILING DATE:	1997-08-22
38	EARLIER	APPLICATION NUMBER:	60/047,595
39	EARLIER	FILING DATE:	1997-05-23
40	EARLIER	APPLICATION NUMBER:	60/047,599
41	EARLIER	FILING DATE:	1997-05-23
42	EARLIER	APPLICATION NUMBER:	60/047,588
43	EARLIER	FILING DATE:	1997-05-23
44	EARLIER	APPLICATION NUMBER:	60/047,585
45	EARLIER	FILING DATE:	1997-05-23
46	EARLIER	APPLICATION NUMBER:	60/047,586
47	EARLIER	FILING DATE:	1997-05-23
48	EARLIER	APPLICATION NUMBER:	60/047,590
49	EARLIER	FILING DATE:	1997-05-23
50	EARLIER	APPLICATION NUMBER:	60/047,594
51	EARLIER	FILING DATE:	1997-05-23
52	EARLIER	APPLICATION NUMBER:	60/047,589
53	EARLIER	FILING DATE:	1997-05-23
54	EARLIER	APPLICATION NUMBER:	60/043,578
55	EARLIER	FILING DATE:	1997-04-11
56	EARLIER	APPLICATION NUMBER:	60/043,576
57	EARLIER	FILING DATE:	1997-04-11
58	EARLIER	APPLICATION NUMBER:	60/047,501
59	EARLIER	FILING DATE:	1997-05-23
60	EARLIER	APPLICATION NUMBER:	60/043,670
61	EARLIER	FILING DATE:	1997-04-11
62	EARLIER	APPLICATION NUMBER:	60/056,632
63	EARLIER	FILING DATE:	1997-08-22
64	EARLIER	APPLICATION NUMBER:	60/056,664
65	EARLIER	FILING DATE:	1997-08-22
66	EARLIER	APPLICATION NUMBER:	60/056,876
67	EARLIER	FILING DATE:	1997-08-22
68	EARLIER	APPLICATION NUMBER:	60/056,881
69	EARLIER	FILING DATE:	1997-08-22

; EARLIER APPLICATION NUMBER: 60/056,909
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,875
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,862
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,887
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/056,908
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/048,964
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/057,650
 ; EARLIER FILING DATE: 1997-09-05
 ; EARLIER APPLICATION NUMBER: 60/056,884
 ; EARLIER FILING DATE: 1997-08-22
 ; EARLIER APPLICATION NUMBER: 60/057,669
 ; EARLIER FILING DATE: 1997-09-05
 ; EARLIER APPLICATION NUMBER: 60/049,610
 ; EARLIER FILING DATE: 1997-06-13
 ; EARLIER APPLICATION NUMBER: 60/061,060
 ; EARLIER FILING DATE: 1997-10-02

Query Match 3.7%; Score 34.8; DB 4; Length 1990;

Best Local Similarity 48.8%; Pred. No. 0.2;

Matches 124; Conservative 0; Mismatches 127; Indels 3; Gaps 1;

QY	592	GGCGAGGAAGTTGTGGTGGGTACAGGTAGTCCGTTCCAGGGAGTTCTTGATGTTGAT	651
DB	871	GGGAGCCACTCAGGTATACGATCCGGTCTCAGCAAGGGCGTTCAGTACGTACGC	930
QY	652	GATTTGGCTGATGCTTGTGTTTTTCTTGCTGGATCGATACAGCGGGTTGGAGCATGTTAAC	711
DB	931	GATCTAGTCAATGGCCTCGTGGCTCTCATGAACAG--CAACGTACAGCAGCCCGGTCAAC	987
QY	712	ATTGGAAGTGGTCAAGAACTACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTT	771
DB	988	CTGGGAACCCAGAGAACACACAAATCTTAGAATTTGCTCAGTTAATTAAGAAACCTTGTT	1047
QY	772	GGTTTTGAAGGAAGCTTGGATGGGATTCGACTTAAGCCAGATGGCACCCGAGGAACCTT	831
DB	1048	GGTAGCGGAAGTGAATTCAGTTCTCTCCGAAGCCAGGATGACCCACAGAAAGAAAA	1107
QY	832	ATGCAGAGCTCAAA	845
DB	1108	CCAGACATCAAAA	1121

RESULT 14

US-09-149-476-131
 ; Sequence 131, Application US/09149476
 ; Patent No. 6420526
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: 186 Human Secreted proteins
 ; FILE REFERENCE: P2002P1
 ; CURRENT APPLICATION NUMBER: US/09/149,476
 ; CURRENT FILING DATE: 1998-09-08
 ; EARLIER APPLICATION NUMBER: PCT/US98/04493
 ; EARLIER FILING DATE: 1998-03-06
 ; EARLIER APPLICATION NUMBER: 60/040,162
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,333
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/038,621
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,626
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,334
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,336
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,163

; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/047,600
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,615
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,597
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,502
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,633
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,583
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,617
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,618
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,503
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,592
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,581
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,584
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,500
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,587
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,492
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,598
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,613
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,582
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,596
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,612
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,632
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,601
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/043,580
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,568
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,314
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,569
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,311
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,671
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,674
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,669
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,312
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,313
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,672
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/043,315
 ; EARLIER FILING DATE: 1997-04-11
 ; EARLIER APPLICATION NUMBER: 60/048,974
 ; EARLIER FILING DATE: 1997-06-06
 ; EARLIER APPLICATION NUMBER: 60/056,886
 ; EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670

EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match 3.7%; Score 34.8; DB 4; Length 2041;
Best Local Similarity 48.8%; Pred. No. 0.2;
Matches 124; Conservative 0; Mismatches 127; Indels 3; Gaps 1;

QY 592 GCGGAGGAAGTTGTGTGGGGTACAGTAGTCCGTTGAGGAGTCTTTCATGTTGAT 651
DB 894 GGGGAGCCACTCAGGTATACGGATCCGGTCTCAGACAAGGGCGTTCCAGTACGTACG 953
QY 652 GATTGGCTGATCTGTGTTTCTTCTGCTGGATCGATACAGCGGTTGGAGCATGTAAC 711
DB 954 GATCTAGTGAATGGCTCTGCTGCTCTCATGAACAG---CAACGTACAGCCCGGTCAAC 1010
QY 712 ATTGGAAGTGGTCAAGAAGTGACTATTAGAGAGTTGGCTCAGTTGCTGAAAGAGTTGTT 771
DB 1011 CTGGGGAACCCAGAGAACAACACATCTTAGAATTTGCTCAGTTAATTAACCTTGT 1070
QY 772 GGTTTTGAAGGAAGCTTGGATGGGATTGCTACTTAAGCCAGATGGCACACCGAGAACTT 831
DB 1071 GGTAGCGGAAGTGAATTCAGTTTCTCTCGAAGCCAGGATGACCCACAGAAAAGAAA 1130
QY 832 ATGGACAGCTCAAA 845
DB 1131 CCAGACATCAAAA 1144

RESULT 15
US-09-738-894A-3/c
Sequence 3, Application US/09738894A
Patent No. 6331423
GENERAL INFORMATION:
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
THEREOF
FILE REFERENCE: CL000636
CURRENT APPLICATION NUMBER: US/09/738,894A
CURRENT FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 36651

```
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(36651)
; OTHER INFORMATION: n = A,T,C or G
US-09-738-894A-3

Query Match      3.4%; Score 32.2; DB 4; Length 36651;
.Best Local Similarity 48.6%; Pred. No. 10;
Matches 88; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 498 TAATCTCTATGGTCCTAATGACAAATTTCCACCCGAGAAATTCATGTGCTTCCTGCTCT 557
Db 35539 TTATATTGTATCCCAATATTTTTTCGCAGCTGTAGAAATTCACCTTGGCTGTGCTTT 35480

Qy 558 TATGAGGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGAAAGTTGTGTGTGGGTAC 617
Db 35479 TGAGAAGTGCTAATTTCTACCATACAGTGTGGCTGGGCAGCCACTTGAATGATGGTTAG 35420

Qy 618 AGGTAGTCCGTTGAGGAGTCTTTCATGTTGATGATTTGGCTGATGCTTTGTTTCTT 677
Db 35419 TATTTTTCATTCAAATTCCTTCTGTTTTTTTGTGTTTGTGTTGTTTTTTTTT 35360

Qy 678 G 678
Db 35359 G 35359
```

Search completed: June 3, 2003, 04:45:32
Job time : 96 secs

PI Nakayama K, Jigami Y;
XX WPI; 2001-381292/40.
DR P-PSDB; AAG62614.
XX
XX Recombinant GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase
PT derived from Arabidopsis for efficient production of GDP (guanine
PT diphosphate)-L-fucose in vitro or in vitro
XX
XX Claim 3; Page 27-29; 35pp; Japanese.
XX
CC The present invention provides the protein and coding sequences of the
CC Arabidopsis thaliana GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-
CC reductase. The sequence can be used for the efficient production of
CC GDP-L-fucose, which is essential to the production of functionally
CC important sugar chains containing L-fucose. The present sequence is the
CC coding sequence of the invention.
XX
SQ Sequence 936 BP; 227 A; 190 C; 240 G; 279 T; 0 other;
Query Match 100.0%; Score 936; DB 22; Length 936;
Best Local Similarity 100.0%; Pred. No. 3.1e-291;
Matches 936; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTGACAAATCTGCCAAATCTTTCGTCGCGGTTCATCGTGGTTGGTTGATCTGCC 60
DB 1 ATGCTGACAAATCTGCCAAATCTTTCGTCGCGGTTCATCGTGGTTGGTTGATCTGCC 60
QY 61 ATTGTCGCGAAGCTTCAGGAACAAGTTTCCAAATCTGTTTCTTAAACACACGCCGAG 120
DB 61 ATTGTCGCGAAGCTTCAGGAACAAGTTTCCAAATCTGTTTCTTAAACACACGCCGAG 120
QY 121 CTTGATCTCACTCGTCAAGCCGATGTTGAATCTCTTTCTCAAGAGAGCCAGTTTAT 180
DB 121 CTTGATCTCACTCGTCAAGCCGATGTTGAATCTCTTTCTCAAGAGAGCCAGTTTAT 180
QY 181 GTAATCCTAGCAGCAGCTAAAGTTGGTGGTATTACCGCTAACACACCTATCCTGCTGAT 240
DB 181 GTAATCCTAGCAGCAGCTAAAGTTGGTGGTATTACCGCTAACACACCTATCCTGCTGAT 240
QY 241 TTCAATGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGACGCGT 300
DB 241 TTCAATGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCATATGACGCGT 300
QY 301 GTGAAGAAGCTTCTTCTCTTGGATCATCTGATTTACCTAAATTTGCTCTCAGCCA 360
DB 301 GTGAAGAAGCTTCTTCTCTTGGATCATCTGATTTACCTAAATTTGCTCTCAGCCA 360
QY 361 ATTCCTGAGTCTCTTTGTTTAAACAGCATCGTCTGAACCACTAATGAGTGGTATGCTATT 420
DB 361 ATTCCTGAGTCTCTTTGTTTAAACAGCATCGTCTGAACCACTAATGAGTGGTATGCTATT 420
QY 421 GCTAAGATCGCTGGGATTAAGCTTGTGAGGCTTATAGGATTCAGACCGATGGGATGCA 480
DB 421 GCTAAGATCGCTGGGATTAAGCTTGTGAGGCTTATAGGATTCAGACCGATGGGATGCA 480
QY 481 ATCTCTGGCATGCTACTAATCTCTATGTCCTTAATGACAATTTCCACCGGAGATTCT 540
DB 481 ATCTCTGGCATGCTACTAATCTCTATGTCCTTAATGACAATTTCCACCGGAGATTCT 540
QY 541 CATGCTCTCTCTCTTATGAGGAGTTTCCAGGCGGAAAGTGAATGAGCGGAGGAA 600
DB 541 CATGCTCTCTCTCTTATGAGGAGTTTCCAGGCGGAAAGTGAATGAGCGGAGGAA 600
QY 601 GTTGTGGTGGGATCAGGTAGTCCGTTGAGGAGTTCTTGCATGTTGATTTGGCT 660
DB 601 GTTGTGGTGGGATCAGGTAGTCCGTTGAGGAGTTCTTGCATGTTGATTTGGCT 660
QY 661 GATGCTTGTGTTTCTTCTGCTGATCGATACAGCGGTTTGAGCATGTTAAACATTTGAAGT 720
DB 661 GATGCTTGTGTTTCTTCTGCTGATCGATACAGCGGTTTGAGCATGTTAAACATTTGAAGT 720
QY 721 GGTCAAGAAAGTACTATTAGAGAGTTGGTGAGTTGGTGAAGAGGTTCTTGGTTTGA 780

DB 721 GGTCAAGAAAGTACTATTAGAGAGTTGGTGAGTTGGTGAAGAGGTTCTTGGTTTGA 780
QY 781 GCGNAGCTTGGATGGATTCGACTAAGCCAGATGCGACACCGAGGAACTTATGGACAGC 840
DB 781 GCGNAGCTTGGATGGATTCGACTAAGCCAGATGCGACACCGAGGAACTTATGGACAGC 840
QY 841 TCAAAGCTCGCTCTTTGGTTGGACACCTAAGGTTTCTCTTAGAGATGCTCTGAGCCAA 900
DB 841 TCAAAGCTCGCTCTTTGGTTGGACACCTAAGGTTTCTCTTAGAGATGCTCTGAGCCAA 900
QY 901 ACTTATGATGGTATTGTAAGAAATGTTTGCACCCGA 936
DB 901 ACTTATGATGGTATTGTAAGAAATGTTTGCACCCGA 936
RESULT 2
AAZ45315
ID AAZ45315 standard; DNA; 1583 BP.
XX
AC AAZ45315;
XX
DT 27-MAR-2000 (first entry)
XX
DE DNA encoding a GDP-4-keto-6-deoxy-D-mannose epimerase/reductase.
XX
KW GDP-4-keto-6-deoxy-D-mannose epimerase/reductase; GDP-D-mannose;
KW GDP-L-galactose; vitamin C; ascorbic acid; L-ascorbic acid;
KW ascorbic acid pathway enzyme; hexokinase; glucose phosphate isomerase;
KW phosphomannose isomerase; phosphomannomutase; L-galactose dehydrogenase;
KW GDP-D-mannose pyrophosphorylase; GDP-D-mannose-GDP-L-galactose epimerase;
KW GDP-L-galactose phosphorylase; L-galactose-1-P-phosphatase;
KW L-galactono-gamma-lactone dehydrogenase; ester; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 49..993
FT /tag= a
FT /product= "GDP-4-keto-6-deoxy-D-mannose epimerase/
FT reductase"
XX
PN WO9964618-A1.
XX
PD 16-DEC-1999.
XX
PF 26-MAY-1999; 99WO-US11576.
XX
PR 08-JUN-1998; 98US-0088549.
PR 17-MAR-1999; 99US-0125073.
PR 18-MAR-1999; 99US-0125054.
XX
PA (DCVB-) DCV INC DBA BIO-TECH RESOURCES.
XX
PI Berry A, Running JA, Severson DK, Burlingame RP;
XX
DR WPI; 2000-105890/09.
DR P-PSDB; AAY54114.
XX
PT Production of ascorbic acid or esters, using microorganisms or plants
PT which have genetic modification in enzymes involved in the ascorbic
PT acid synthesis pathway
XX
PS Claim 26; Page 164-166; 187pp; English.
XX
CC The present sequence encodes a GDP-4-keto-6-deoxy-D-mannose epimerase/
CC reductase. The enzyme catalyses the conversion of GDP-D-mannose to
CC GDP-L-galactose. The enzyme can be modified, and used to produce
CC transgenic microorganisms, which can be used in fermentation techniques
CC to produce vitamin C (ascorbic acid, L-ascorbic acid). The enzyme is
CC modified to increase its action. Other ascorbic acid pathway enzymes
CC which may be used in the method of the invention include hexokinases,
CC glucose phosphate isomerases, phosphomannose isomerases,

CC phosphonammutases, GDP-D-mannose pyrophosphorylases,
CC GDP-D-mannose:GDP-L-galactose epimerases, GDP-L-galactose phosphorylases,
CC L-galactose-1-P-phosphatases, L-galactose dehydrogenases, and
CC L-galactono-gamma-lactone dehydrogenases. The methods can be used for
CC the production of ascorbic acid or esters using microorganisms or plants.
XX
SQ Sequence 1583 BP; 407 A; 330 C; 351 G; 495 T; 0 other;

Query Match 91.6%; Score 857; DB 21; Length 1583;
Best Local Similarity 97.1%; Pred. NO. 1.3e-265;
Matches 918; Conservative 0; Mismatches 15; Indels 12; Gaps 4;
QY 1 ATGTCGACAAATCTGCCAAATCTTCGTCGGGTCATCGTGTGGTGGTGAATCGC 60
DB 49 ATGTCGACAAATCTGCCAAATCTTCGTCGGGTCATCGTGTGGTGGTGAATCGC 108
QY 61 ATGTCGCGAAGCTTCAGGAACAAGGTTTCCAAATCTCTGTTCTTAAACACACGCGGAG 120
DB 109 ATGTCGCGAAGCTTCAGGAACAAGGTTTCCAAATCTCTGTTCTTAAACACACGCGGAG 168
QY 121 CTGATCTCTACTCGTCAAGCCGATGTTGAATCCTCTTTCTCAAGAGAAGCCAGTTAT 180
DB 169 CTGATCTCTACTCGTCAAGCCGATGTTGAATCCTCTTTCTCAAGAGAAGCCAGTTAT 228
QY 181 GTAATCTAGCAGCAGCTAAAGTTGTGTGTATTCAGCTAAACACCTATCTGCTGAT 240
DB 229 GTAATCTAGCAGCAGCTAAAGTTGTGTGTATTCAGCTAAACACCTATCTGCTGAT 288
QY 241 TTCTATTGGTGTCAATCTCAGATTTCAGACCAATGTGATCCACTTCGCATATGAGCAGCGT 300
DB 289 TTCTATTGGTGTCAATCTCAGATTTCAGACCAATGTGATCCACTTCGCATATGAGCAGCGT 348
QY 301 GTGAAGAAGCTTCTCTCTCTGATCATCTGATCATCTGATTAACCTTAAATTTCTCTCAGCCA 360
DB 349 GTGAAGAAGCTTCTCTCTCTGATCATCTGATCATCTGATTAACCTTAAATTTCTCTCAGCCA 408
QY 361 ATTCCTGATCTGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 409 ATTCCTGATCTGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 468
QY 421 GCTAAGATCGTGGGATTAAGACTTCTGAGGCTTATAGGATTCAGCAGCGATGGATGCA 480
DB 469 GCTAAGATCGTGGGATTAAGACTTCTGAGGCTTATAGGATTCAGCAGCGATGGATGCA 528
QY 481 ATCTCTGGATGCTACTAATCTCTATGCTCTTAATGACAAATTTCCACCGGAGATTTCT 540
DB 529 ATCTCTGGATGCTACTAATCTCTATGCTCTTAATGACAAATTTCCACCGGAGATTTCT 585
QY 541 CATGCTCTCTGCTCTTATGAGGAGGTTCCAGGAGCGGAAAGTGAA-TGGAGCGGAGGA 599
DB 586 CATGCTCTCTGCTCTTATGAGGAGGTTCCAGGAGCGGAAAGTGAAATGGAGCGGAGGA 645
QY 600 AGTTGTGGTGTGGGTACAGGATGTC- - - - -GTTGAGGAGGTTCTGATGATGAT 654
DB 646 AGTTGTGGTGTGGGTACAGGATGTC- - - - -GTTGAGGAGGTTCTGATGATGAT 705
QY 655 TTGGCTGATGCTGTGTTTCTTCTGCTGATCG- - - - -ATACAGCGGTTTGAGGATGTTTAC 711
DB 706 TTGGCTGATGCTGTGTTTCTTCTGCTGATCG- - - - -ATACAGCGGTTTGAGGATGTTTAC 765
QY 712 ATGGAAGTGGTCAAGAAGTACTATAGAGAGTGGCTGAGTTGGTGAAGAGGTTGTT 771
DB 766 ATGGAAGTGGTCAAGAAGTACTATAGAGAGTGGCTGAGTTGGTGAAGAGGTTGTT 825
QY 772 GTTTTTGAAGGAAGCTTCGATGGATTCACCTAAGCAGATGGCACCGGAGAACTT 831
DB 826 GTTTTTGAAGGAAGCTTCGATGGATTCACCTAAGCAGATGGCACCGGAGAACTT 885
QY 832 ATGGACAGCTCAAAGCTCCGCTCTTGGGTGGTGGACACCTAAGGTTTCTCTTAGAGATGTT 891
DB 886 ATGGACAGCTCAAAGCTCCGCTCTTGGGTGGTGGACACCTAAGGTTTCTCTTAGAGATGTT 945
QY 892 CTGAGCCAACTTATGATGTTGTTTGAAGAATGTTTGAACCGA 936

Db 946 CTGAGCCAACTTATGATGTTTGAAGAATGTTTGAACCGA 990
RESULT 3
AAC34827
ID AAC34827 standard; DNA; 1353 BP.
XX AAC34827;
AC AAC34827;
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 8033.
DE Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

```
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 70.4%; Score 659; DB 21; Length 1353;
Best Local Similarity 82.6%; Pred. No. 9.1e-202;
Matches 766; Conservative 0; Mismatches 160; Indels 1; Gaps 1;

QY 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGTCATCTGTTGGTTGGATCTGCC 60
    |||||
Db 199 ATGTTAGAGAAATCCGCCAAATCTTTTCTGTCGCCATAGGGGATTTGGTTCGCC 258
    |||||
QY 61 ATTGTCCGCAAGCTTCAGGAACAAGTTTTCACCAATCTCGTTCTTTAAACACACGCCGAG 120
    |||||
Db 259 ATATCCCGCAAGCTTCAGGATCAGGGTTTTCACCAATCTCGTTCTTCGAAACATTCGAG 318
    |||||
```


Db 704 GCGCGAACTTCTGATGTCAGGATCTGCGGAGCGCATGCTGTCTCTTTTGGAACTT 763
 QY 689 ACAGCGGGTTGGAGCATGTTAAACATTTGGAAGTGGTCAAGAAGTGAATTTAGAGAGTTGG 748
 Db 764 TCGATGGTCCGAACCAACGTCACAGTGGGACCGGCTCGATCACAGCATTTAGCGAGATCG 823
 QY 749 CTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGGAAGCTTGGATGGGATTCACATAAGC 808
 Db 824 CAGACATGGTCTGCTACGGCGGTGGGCTACATCGCGGAAACACAGTTTGGGATCCAACTAAAC 883
 QY 809 CAGATGGCACACCGAGGAACCTTATGGACAGCTCAAGAGCTCGCTCTTTGGGTTGGACAC 868
 Db 884 CCGATGGAAACCCGCGCAAACTATTGGACCTCTCCGCGTACGCGAGTTGGGTTGGCGCC 943
 QY 869 CTAAGTTTCTCTTAGAGATGGTCTGAGCCAACTTATGATGGTATTTGAAGAATG 925
 Db 944 CCGCAATCGCACTGAAAGAGCGCATCGATGCAACGGTGTGCTGATACCGCACAAATG 1000

RESULT 6
 ID AAT74479
 XX AAT74479 standard; DNA; 4435 BP.
 AC AAT74479;
 XX
 DT 11-MAR-1998 (first entry)
 XX
 DE Part of the GS region DNA sequence from M. paratuberculosis.
 XX
 KW GS; pathogenicity island; pathogenic protein; mycobacterial disease;
 KW cellular recognition receptor; pathogenic mycobacteria; Crohn's disease;
 KW vaccine; inflammatory disease; sarcoidosis; Johne's disease; ss.
 XX
 OS Mycobacterium paratuberculosis.
 XX
 FH Key Location/Qualifiers
 FT CDS 201..1232
 FT /tag= a
 FT /note= "encodes AAW21769"
 FT CDS 1172..2191
 FT /tag= b
 FT /note= "encodes AAW21771"
 FT CDS 2467..3189
 FT /tag= c
 FT /note= "encodes AAW21773"
 FT CDS 3335..4135
 FT /tag= d
 FT /note= "encodes AAW21775"
 XX
 PN WO9723624-A2.
 XX
 PD 03-JUL-1997.
 XX
 PF 23-DEC-1996; 96WO-GB03221.
 XX
 PR 21-DEC-1995; 95GB-0026178.
 XX
 PA (SGEO-) ST GEORGE'S HOSPITAL MEDICAL SCHOOL.
 XX
 PI Doran T, Ford J, Hermon-Taylor J, Loughlin M, Millar D;
 PI Sumar N, Tizard M;
 XX
 DR WPI; 1997-351061/32.
 XX
 PT New isolated pathogenicity island from mycobacteria - used to
 PT develop products for detection, diagnosis, prevention and treatment
 PT of mycobacteria infections
 XX
 PS Claim 5; Pages 43-44; 62pp; English.
 XX
 CC The present sequence represents a novel polynucleotide sequence
 CC designated "GS". GS is a pathogenicity island of 8 kb of DNA comprising
 CC a core region of 5.75 kb with multiple open reading frames (ORFs) and

CC an adjacent transmissible element of 2.5 kb. The ORFs, and also the
 CC transmissible element, encode proteins which may be linked to
 CC pathogenicity, such as providing receptors for cellular recognition. GS
 CC was discovered and characterised using differential DNA analysis
 CC technology. It is found within Mycobacterium paratuberculosis and it has
 CC also been identified in Mycobacterium avium subspecies silvaticum. These
 CC pathogenic mycobacteria cause chronic inflammation of the intestine and
 CC Crohn's disease in humans. The protein products of the ORFs of GS can be
 CC used for detecting mycobacteria or for diagnosing, treating or preventing
 CC mycobacterial diseases. In particular they can be used as vaccines for
 CC inflammatory diseases such as Crohn's disease or sarcoidosis in humans or
 CC Johne's disease in animals.
 XX

Sequence 4435 BP; 904 A; 1289 C; 1288 G; 954 T; 0 other;
 Query Match 31.7%; Score 297; DB 18; Length 4435;
 Best Local Similarity 58.2%; Pred. No. 1.1e-84;
 Matches 522; Conservative 0; Mismatches 375; Indels 0; Gaps 0;

QY 29 TCGCGGTCATCGTGGTTGGTTCGCAATTCGCGAAGCTTCAGGAACAAGGTT 88
 Db 1275 TCGCGGTCATCGGGGGTGGTGGCTCAGCGCTCGTAGATTGAGGCGCGAGGGT 1334
 QY 89 TCACCAATCTCGTTCTTAAACACACGCGGAGCTTGATCTCACTCGTCAAGCGGATGTG 148
 Db 1335 TCACCAATCTCATTTGTGCGATCACGCGATGAGATTGATCTGACGACCGGCGCAACGT 1394
 QY 149 AATCCTCTTTTCTCAAGAGAGCCAGTTTATGTAATCTAGCAGCAGCTAAAGTTGGTG 208
 Db 1395 TTGATTTGTGTCGAGCAACACACAGGTGATCATGATCGCGCGGACCGGTCGGCG 1454
 QY 209 GTATTACGCTAACACACCTATCTCTGCTGATTTCATTGGTGTCAATCTCCAGATTCAGA 268
 Db 1455 GCATCATGGCGAATAACACCTATCCGGGAGCTTCTTGTCGAAACCTCCGAATCCAGA 1514
 QY 269 CCAATGTGATCCACTCTGCATATGACACGGTGTGAAGAGCTTCTTCTCTTGGATCAT 328
 Db 1515 CCAATTTGCTCGACGAGCTGTCGCGGTGTCGCGGCTCTCTTCTCGGTTCGT 1574
 QY 329 CCGTCAATTTACCTAAATTTGCTCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCAT 388
 Db 1575 CATGATCTACCCGAACTAGCTCCGCAACCTATCCAGAGAGTCTTTATTGATGGCC 1634
 QY 389 CGCTTGAACCAACTAATGAGTGGTATGCTATTGTGAAGATCGCTGGGATTAAGACTTGTG 448
 Db 1635 CTTTGGAGCCCAACCAACGCGTATCGATCGCAAGATCGCGTATCCTGCAAGTTC 1694
 QY 449 AGGCTTATAGGATTCAGCACGGATGGGATGCAATCTCTGGGATGCCCTACTATCTCTATG 508
 Db 1695 AGGCGGTTAGCGCGCAATATGGGCTGGGCTGGATCTCTGCGATGCGGACTTAACCTCTACG 1754
 QY 509 GTCTTAATGACAATTTCCACCGGAGAAATTCCTCATGTGCTTCTCTTATGAGGAGGT 568
 Db 1755 GACCGGCGACAATTTCTCCCGTCCGGTCCGATCTCTTTCGCGGCTCATCCGTCGAT 1814
 QY 569 TCCACGAGCGCAAAAGTGAATGGAGCGGAGGAAATTTGTGTGGGTGACAGGTAGTCCGT 628
 Db 1815 ATGAGGAAGCCAAAGCTGGTGGTCAGAAAGAGTGAAGATGGGGGACCGGTACTCCGC 1874
 QY 629 TGAGGGAGTTCTTGCAATGTTGATGATTTGGCTGATGCTTGTGTTTCTTCTGCTGATCGAT 688
 Db 1875 GCGCGGAACTTCTGCAATGTCGACGATCTGGCGAGCGCATGCTCTCTTTTGGAACTT 1934
 QY 689 ACAGCGGGTTGGAGCATGTTTAACTTGAAGTGGTCAAGAACTGACTATTAGAGAGTTGG 748
 Db 1935 TCGATGGTCCGAAACCAACGTCACGTCGCGGCGATCCAGGCGTGCATCAGCATAGCAGATCG 1994
 QY 749 CTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGGAAGCTTGGATGGGATTCGACTTAAGC 808
 Db 1995 CAGACATGGTCTAGCGGGTGGCTACATCGCGCAACACGTTGGGATCCAACTAAAC 2054
 QY 809 CAGATGGCACACCGAGGAACCTTATGGACAGCTCAAGAGCTCAAGAGCTTGGGTTGGACAC 868

Db	1704180	ACGCACGTTTGGCGGCGGGTTCACCAACCTGCTGGTGGTCAACGCGCCGAGCTGA	1704239	XX	Escherichia coli.
QY	126	TCTCACTCGTCAAGCGGATGTTGAATCTCTTTCTCAAGAGAAGCCAGTTTATGTAAT	185	XX	Key
Db	1704240	TCTGACGGATCGGCGCGGAGTTCGACTTCGTTCTCGAGTCGAGCGCGAGTCGTCAAT	1704299	FT	CDS
QY	186	CCTAGCAGCAGCTAAAGTTGGTGGTATTCACGCTAAACAACCATCTCTGCTGATTTTCAT	245	FT	Location/Qualifiers
Db	1704300	CGACGCGGCGCGGGTCCGGCGCATCTCTGGCCACGACACCTACCGGCCGATTTCTT	1704359	FT	79..861
QY	246	TGCTGTCAATCTCCAGATTCAGACCAATGATGATCCACTCTGTCATATGAGCACCGTGTGAA	305	FT	/*tag= a
Db	1704360	GTCGGAAAACCTCCAGATCCAGTCAACCTCTGCTGGATCGCGCGTGGCGCGGGTGC	1704419	FT	/product= sugar transferase
QY	306	GAAGCTTCTCTCTCTGGATCATCTGCAATTTACCTTAATTTGCTCCTCAGCCATTC	365	FT	/note= "wbdN gene (ORF1), this region is specifically claimed in Claim 6; encodes AAW88312"
Db	1704420	CGCGTGTCTTCTCTGGCTCTGCTGCTGATCTACCCGAACTCGCCCGGAGCGATCCC	1704479	FT	858..2042
QY	366	TGAGTCTGCTTTGTTAAACAGCATCGCTTGAACCAACTAATAGTGGTATGCTATTGCTAA	425	FT	/*tag= b
Db	1704480	GGAGAGCGCGCTCTCACCGTCCGTTGGAGCCGACCAACGACGCGTACGCGATCGCCAA	1704539	FT	/product= O antigen-polymerase
QY	426	GATCGCTGGGATTAAGACTTGTGAGGTTATAGGATTCAGACGGATGGGATGCAATCTC	485	FT	/note= "wzy gene (ORF2), this region is specifically claimed in Claim 6; encodes AAW88313"
Db	1704540	AATCGCGCGCATCTTTCGGTCCAGCGGTGCGCCGCAACATGGCTGCGGTGGATCTC	1704599	FT	2011..2757
QY	486	TGCGATGCTCTACTTAATCTCTATGCTCTAATGACAAATTTCCACCCGAGAAATTTCTCATCT	545	FT	/*tag= c
Db	1704600	GGCGATGCCCAACACTGTACGGGCGAGCGCAAACTTTTCGCGTCCGGTCTCGCATCT	1704659	FT	/product= sugar transferase
QY	546	GCTTCTGCTCTTATGAGGAGGTTTCCACGAGGCGAAAGTGAATGGAGCGGAGAACTTGT	605	FT	/note= "wbdO gene (ORF3), this region is specifically claimed in Claim 6; encodes AAW88314"
Db	1704660	GCTGCGGCACTCATCCGCGCTATGACGAGGCCAAGCCAGTGGCGCGCCCAACGTGAC	1704719	FT	2744..4135
QY	606	GGTGTGGGTACAGGTAGTCCGTTGAGGGAGTTCTTGCATGTTGATGATTTGCTGATGC	665	FT	/*tag= d
Db	1704720	CAACTGGGGCACCGGCACCGCCCGAGGGAGTTGCTGCACGTCGACGACCTGGCGAGCGC	1704779	FT	/product= O antigen flippase
QY	666	TTGCTGTTTCTCTCGATCGATACAGCGGTTGGAGCATGTTAACTTGAAGTGTCA	725	FT	/note= "wzx gene (ORF4), this region is specifically claimed in Claim 6; encodes AAW88315"
Db	1704780	ATCCCTGTATCTCTCGGAACATTTCCAGCGGCGGACCCCATGTCAACGTGGGAACCGGCAT	1704839	FT	4132..5232
QY	726	AGAAGTGACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTTGGTTTGAAGGAA	785	FT	/*tag= e
Db	1704840	CGACCAACCATCGCGAGATCCCGAGATGGTCCGCTCGCGGTAGGCTATAGCGGCA	1704899	FT	/product= perosamine synthetase
QY	786	GCTTGGATGGATTCACCTAAGCAGATGCGACACCGAGGAACTTATGACAGCTCAAA	845	FT	/note= "per gene (ORF5); encodes AAW88316"
Db	1704900	AACCCGCTGGATTCGAAGCAAAACCGGACGGAACACCAACGAACTGCTGGATGTTTCGGT	1704959	FT	5257..6471
QY	846	GCTCGGCTCTTTGGGTTGACACCTAAGGTTTCTCTTAGAGATGGTCTGAGCCAACTTA	905	FT	/*tag= f
Db	1704960	GCTACGGAGGCGGATGGCGGCTTCGATCGCGCTGCGGACGCGATCGAGCGGACGCT	1705019	FT	/product= sugar transferase
QY	906	TGATTTGGTATTTGAAGAATG 925		FT	/note= "wbdP gene (ORF6), this region is specifically claimed in Claim 6; encodes AAW88317"
Db	1705020	GGCGTGTATCGGAGCAGC 1705039		FT	6491..7609
RESULT 12				FT	/*tag= g
AA06749				FT	/note= "ORF7; encodes AAW88318"
ID	AA06749	standard; DNA; 14024 BP.		FT	7606..8578
XX	AA06749;			FT	/*tag= h
AC				FT	/note= "ORF8; encodes AAW88319"
XX				FT	8580..9089
XX	26-APR-1999	(first entry)		FT	/*tag= i
DE	E. coli O157 antigen gene cluster.			FT	/note= "ORF9; encodes AAW88320"
DE	O antigen; O157 antigen; diagnosis; wbdN gene; wzx gene; wzy gene;			FT	9071..10519
KW	wbdP gene; wbdO gene; wbdR gene; transferase; polymerase; flippase;			FT	/*tag= j
KW	diarrhoea; haemorrhagic colitis; ds.			FT	/note= "manC gene (ORF10); encodes AAW88321"
				FT	10523..11893
				FT	/*tag= k
				FT	/note= "manD gene (ORF11); encodes AAW88322"
				FT	12007..13123
				FT	/*tag= l
				FT	/note= "H-repeat, may play a role in gene cluster assembly"
				FT	13156..13821
				FT	/*tag= m
				FT	/product= N-acetyl transferase
				FT	/note= "wbdR gene (ORF12), this region is specifically claimed in Claim 6; encodes AAW88323"
				XX	
PN	WO9850531-A1.			XX	
XX	12-NOV-1998.			XX	
PD				XX	
XX	01-MAY-1998;			XX	98WO-AU00315.
XX				XX	
PR	22-JUL-1997;			XX	
XX	01-MAY-1997;			XX	97AU-0008162.
				XX	97AU-0008545.

(UNSY) UNIV SYDNEY.

Reeves PR, Wang L;

WPI; 1999-059669/05.
P-PSDB; AAW88312, AAW88313, AAW88314, AAW88315, AAW88316, AAW88317,
AAW88318, AAW88319, AAW88320, AAW88321, AAW88322, AAW88323.

Nucleic acid molecules specific for bacterial polysaccharide
antigens - useful for detecting specific strains in, e.g. food,
faeces or patient samples

Claim 6; Fig 8; 165pp; English.

This is the nucleotide sequence of a gene cluster involved in
the biosynthesis of the *Escherichia coli* O157 O antigen. The
gene cluster was obtained by PCR amplification of *E. coli*
O157:H7 strain (C664-1992) DNA using primers (see AAX06752-63) based
on O antigen gene sequences. 12 Open reading frames within the
gene cluster were identified on the basis of homology to known
sequences. The genes encode proteins (see AAW88312-23) involved in the
synthesis of sugars present in the polysaccharide antigen, and in
the transport or processing of polysaccharide or oligosaccharide
units. The use of nucleic acid molecules derived from particular
assembly and transport genes, particularly wbd (transferase), wzx
(flippase) and wzy (polymerase) genes, within O antigen gene
clusters improves the specificity of methods for the detection and
identification of O antigens, e.g. in tests of food- or faecal-
derived samples, or samples from patients. The O antigen is a
major virulence factor of enteropathogenic *E. coli* strains that
cause diarrhoea and haemorrhagic colitis.

Sequence 14024 BP; 4546 A; 2263 C; 2832 G; 4383 T; 0 other;

Query Match 28.8%; Score 269.8; DB 20; Length 14024;

Best Local Similarity 57.1%; Pred. No. 1.2e-75;

Matches 530; Conservative 0; Mismatches 372; Indels 27; Gaps 1;

22 ATCTTCGTCGCGGTCATCGTGGTGGTGGATCTGCCAATGTCGCAAGCTTCAGGAA 81

Db 7627 ATTTTATTGCTGGTCAACAAGGATGGTTGGATCAGTATTACCGGCGCTCAACAA 7686

Qy 82 CAAGGTTTCAACCATCTGTTCTTAAACACACGCGAGCTTGATCTCACTCGTCAGCC 141

Db 7687 CGTGATGATGTTGAGTGGTGTGTACGTACTCGGGATGAATTGAATGTTGGATAGC 7746

Qy 142 GATGTTGAATCCTCTTTCTCAAGAGAACCCAGTTTATGTAATCCTAGCAGCAGTAA 201

Db 7747 GCTGTTTGGATTTTTTTTCTTACAGAAATGACACCGAGTTTATTTGGCAGCAGAAA 7806

Qy 202 GTTGGTGGTATTACGCTAACACACCTATCTGCTGATTTTCATTGGTGTCAATCTCCAG 261

Db 7807 GTCGGAGGTATTAGTCAACAGTTCTTATCTGCGGATTTATATATAGATATAAGT 7866

Qy 262 ATTCAGACCAATGTGATCACTCTGATATGACGCGGTGTGAAGAGCTTCTCTTCCTT 321

Db 7867 ATAGAGCGAATGTCAATCATGTCGCCCAAAAATAATGTAATAAACTGCTTTTCTCTC 7926

Qy 322 GGATCATCTCGATTTACCTTAATTTGCTCTCAGCAATTCCTGAGTCTGCTGTGTA 381

Db 7927 GGTTCGTCGTGTTATTTATCTTAAGTTAGCACCAACCGATTTAGGAAGCAAGATTATTA 7986

Qy 382 ACAGCATCTGTTGAACCAACTAATGATGTTGATGCTATTGCTAAGATCGCTGGATTAAG 441

Db 7987 CAAGGAAATCTGAGCAACAATGAACCTTATGCTATCGCAAAATTCAGGATTATAA 8046

Qy 442 ACTGTGAGGTTATAGGATTCAGCAGGATGGATGCAATCTCTGGATGCTCACTAAT 501

Db 8047 TTATGTGAATCTTATAACCGTCAGTTTGGCGGTGATTTACCGTTTCAGTAATGCAACCAAT 8106

Qy 502 CTCTATGTCCTAATGACAAATTCACCGGAGAAATTCATGCTGCTCTCTCTTATG 561

Db 8107 CTTTATGGTCCAAATGACAAATTTTCAATCAAGTAATTCATGTAATCCGGCGCTTTTG 8166

Qy 562 AGGAGGTTCCACGAGCGGAAAGTGAATGGAGCGGAGGAAGTTGTGTGTGGGTACAGGT 621

Db 8167 CGCCGCTTTTCATGATGCTGTGGAAAACAATTTCTCCGAATGTTGTTGTTGGGAAAGTGGT 8226

Qy 622 AGTCGCTTTCAGGAGGATCTTTCATGATGATTTGGCTGATGCTGTTGTTTCTTCTGCTG 681

Db 8227 ACTCCAAAGCGTGAATTTTACATGTAGATGATATGGCTTCTGCAAGCAATTTATGTCATG 8286

Qy 682 G-----ATCGATACAGCGGTTGGAGCATGTTAAACATT 714

Db 8287 GAGATGCCATACATATATGCGAAAATACTAAAGTAATGTTGTTCTATATCAATATT 8346

Qy 715 GGAAGTGTCAAGAAGTCTATTAGAGAGTTGGTGTGAGTTGGTGAAGAGGTTGTTGGT 774

Db 8347 GGAACAGGTTATTGACTGTCACGATTTTGTGAGCTTTCGGAACAATAGCAAAAGTTGTAGGT 8406

Qy 775 TTTGAGGGAAGCTTGGATGGATTCACCTAAGCCAGATGGCACACCGAGGAACCTTATG 834

Db 8407 TATAAAGGCGATATTACGTTTCGATACAAACAAAGCCGATGGAGCCCTCGAAAACCTACTT 8466

Qy 835 GACAGCTCAAGCTCGCTTTTGGGTTGGACACCTAAGGTTTCTCTTAGAGATGCTCTG 894

Db 8467 GATGTAACGCTTCTTCATCACTAGTTGGATCATAAATTAACCTTCAAGGCTCTT 8526

Qy 895 AGCCAAACTTATGATTTGTTGAAGAA 923

Db 8527 GAAATACATACAACTGTTTCTTGAAGAA 8555

RESULT 13

AAZ56386

ID AAZ56386 standard; DNA; 14024 BP.

XX AC AAZ56386;

XX DT 17-MAR-2000 (first entry)

XX DE *Escherichia coli* flagellin protein nucleotide sequence SEQ ID NO:56.

XX KW Flagellin; fliC; antigen; detection; ds.

XX OS *Escherichia coli*.

XX PN WO961458-A1.

XX PD 02-DEC-1999.

XX PF 21-MAY-1999; 99WO-AU00385.

XX PR 21-MAY-1999; 98AU-0003634.

XX PA (UNSY) UNIV SYDNEY.

XX PI Reeves PR, Wang L;

XX DR WPI; 2000-072598/06.

XX PT Novel nucleic acid molecule useful for the detection of flagellated

XX PS Claim 3; Page 229-233; 245pp; English.

XX CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all

or part of an *Escherichia coli* flagellin protein except a protein

expressed by *E. coli* H1, H7, H12 or H48 type strains. The present

invention also describes a method of detecting the presence of *E. coli*

of a particular H serotype in a sample, comprising specifically

hybridising a nucleic acid, preferably at least a pair, derived from a

flagellating gene, specific for a particular flagellin gene associated

with the H serotype, to any *E. coli* in the sample which contain the gene,

and detecting any hybridised molecules, identifying the presence of that

serotype in the sample. (I) are useful for: (1) detecting the presence

FT CDS /note= "homologous to the Tnp gene"
FT complement (433880..434110)
FT *tag= k
FT /standard_name= "ORF K11"
FT /product= "protein of unknown function"
FT /note= "homologous to the FixU gene"
FT complement (434107..434433)
FT *tag= l
FT /standard_name= "ORF K12"
FT /product= "protein of unknown function"
FT complement (434517..434711)
FT *tag= m
FT /standard_name= "ORF K13"
FT /product= "ferrodoxin/ferrodoxin-like protein"
FT /note= "homologous to the FdxN gene"
FT complement (434753..436234)
FT *tag= n
FT /standard_name= "ORF K14"
FT /gene= "nifB"
FT /product= "protein involved in FeMo co-factor
FT biosynthesis"
FT complement (436460..438130)
FT *tag= o
FT /standard_name= "ORF K15"
FT /gene= "nifA"
FT /product= "positive regulator of nif, fix and other
FT genes"
FT complement (438297..438590)
FT *tag= p
FT /standard_name= "ORF K16"
FT /gene= "fixX"
FT /product= "protein required for nitrogenase activity"
FT complement (438605..439912)
FT *tag= q
FT /standard_name= "ORF K17"
FT /gene= "fixC"
FT /product= "protein required for nitrogenase activity"
FT complement (439923..441032)
FT *tag= r
FT /standard_name= "ORF K18"
FT /gene= "fixB"
FT /product= "protein required for nitrogenase activity"
FT complement (441042..441899)
FT *tag= s
FT /standard_name= "ORF K19"
FT /gene= "fixA"
FT /product= "protein required for nitrogenase activity"
FT complement (442316..442636)
FT *tag= t
FT /standard_name= "ORF K20"
FT /product= "protein of unknown function"
FT complement (443313..443879)
FT *tag= u
FT /standard_name= "ORF K21"
FT /product= "protein of unknown function"
FT 444337..445029
FT *tag= v
FT /standard_name= "ORF K22"
FT /product= "ferrodoxin-like protein"
FT /note= "homologous to the NifQ gene"
FT 445088..446602
FT *tag= w
FT /standard_name= "ORF K23"
FT /gene= "dctA"
FT /product= "C4-dicarboxylate transport protein"
FT /note= "homologous to the DctAI gene"
FT 446599..447843
FT *tag= x
FT /standard_name= "ORF L1"
FT /product= "cytochrome P450-like protein"
FT /note= "homologous to the CamC gene"
FT 447844..448500
FT *tag= y

FT FT /standard_name= "ORF L2"
FT /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT protein"
FT /note= "homologous to the LinA gene"
FT 448497..450203
FT *tag= z
FT /standard_name= "ORF L3"
FT /product= "putative protein with degradative function"
FT 450341..451396
FT *tag= aa
FT /standard_name= "ORF L4"
FT /product= "luciferase alpha-subunit-like protein"
FT /note= "homologous to the LuxA gene"
FT 452980..454494
FT *tag= ab
FT /standard_name= "ORF L6"
FT /gene= "nifD"
FT /product= "alpha-subunit of FeMo protein of nitrogenase"
FT 454590..456131
FT *tag= ac
FT /standard_name= "ORF L7"
FT /gene= "nifK"
FT /product= "beta-subunit of FeMo protein of nitrogenase"
FT 456187..457677
FT *tag= ad
FT /standard_name= "ORF L8"
FT /product= "protein involved in FeMo co-factor
FT biosynthesis"
FT /note= "homologous to the NifB gene"
FT 457887..459096
FT *tag= ae
FT /standard_name= "ORF L9"
FT /product= "protein involved in FeMo co-factor
FT biosynthesis"
FT /note= "homologous to the FixF gene"
FT 459093..459575
FT *tag= af
FT /standard_name= "ORF L10"
FT /product= "protein of unknown function"
FT /note= "homologous to the NifX gene"
FT 459579..460067
FT *tag= ag
FT /standard_name= "ORF L11"
FT /product= "protein of unknown function"
FT 460501..460920
FT *tag= ah
FT /standard_name= "ORF L12"
FT /product= "protein similar to part of the Fe protein
FT of nitrogenase"
FT /note= "homologous to the NifH gene"
FT 461228..461545
FT *tag= ai
FT /standard_name= "ORF L13"
FT /product= "protein of unknown function"
FT 463201..464739
FT *tag= aj
FT /standard_name= "ORF L14"
FT /product= "peptidase-like protein"
FT /note= "homologous to the bi-MPP gene"
FT 464736..466079
FT *tag= ak
FT /standard_name= "ORF L15"
FT /product= "processing protease-like protein"
FT /note= "homologous to the PP gene"
FT 466590..467021

Query Match 28.3%; Score 265; DB 19; Length 534720;
Best Local Similarity 56.6%; Pred. No. 3e-73;
Matches 512; Conservative 0; Mismatches 390; Indels 3; Gaps 1;
QY 22 ATCTTCGTCGGGTCATCGTGGTTGGTTGGATCTGCATGTCGCAAGTTCAGNA 81
DB 6087 ATTTGGTTCGAGGACACAAAGTATGTCGCGCAGCCATAATTGATCGTTCCTCC 6028

QY 82 CAAGGTTTACCAATCTCGTCTTAAACACACGCGAGCTTGATCTCACTCGTCAAGCC 141
 DB 6027 GAGGATTGGAAGTCACTGTCAGATAGCA---AAAGCTTGATCTGACGCGCAAGAG 5971
 QY 142 GATGTTGAATCTCTTTTCTCAAGAGAACGAGTTTATGTAATCCTAGCAGCAGCTAAA 201
 DB 5970 GAAGTTGAGAAATTTCTATTAAAGGAAAGCCGACGCGGTCTAATATGCGCGCAGGAAG 5911
 QY 202 GTTGGTGGTATTACAGCTTAACACACCTATCTGCTGATTTTCAATGGTGTCAATCTCCAG 261
 DB 5910 GTGGGTGGGATCTCGCAATGATATGATGCGCGCTGACTTCTATCAAAAACCTTATC 5851
 QY 262 ATTACAGACCAATGTGATCACTCTGCATATGACACGCGTGTGAAGAGCTTCTCTTCTTT 321
 DB 5850 ATGGAGGCTATGTCATTGAGGGCTCTTCGCGAGTGGCGTTGAAGAGCTTCTTCTTCT 5791
 QY 322 GGATCATCTGCAATTTACCTAAATTTGCTCTCAGCCAAATCTCTGAGTCTCTTTGTTA 381
 DB 5790 GGATCGAGTTGCATATATCGAAGTATGCGCGCAGCCCAATAAGGAAGAGGCTCTATTA 5731
 QY 382 ACAGATCTGCTGAACCACTAATGAGTGTATGCTATTTGCTTAAGATCGCTGGATTAAG 441
 DB 5730 ACCGGACCACTTGAGCGGACCAACGAGTGTGATGCGATCGCCAAATCGCGGCATTAA 5671
 QY 442 ACTGTGACGCTTATAGGATTCAGACGCGATGGATGCAATCTCTGGCATGCTCTACTAAT 501
 DB 5670 TTGTGTCAGCGTATCTGAAGCAATACGCGGCAACTTCAATATAGCCATGCGCAAT 5611
 QY 502 CTCTATGGTCTTAATGACAAATTTCCACCGGAGAAATTCATGATGCTCTCTCTCTATG 561
 DB 5610 CTCTATGGCCACCGCATTAAGTTCGATCTTAATCTCCAGCCAGCTCGTCTCTCTCTAATA 5551
 QY 562 AGGAGTTTCCAGGCGGAAAGTGAATGAGCGGAGGAAGTTGTGGTGTGGGTACAGT 621
 DB 5550 CGCAAGACATGAGGCAAGATTAAGACCTTGGGTGCTTGTCTATATATGGGAAAGCGC 5491
 QY 622 AGTCCGTTGAGGAGTCTTGATGTTGATGTTGGCTGATGCTTGTGTTTCTTGTG 681
 DB 5490 ACACCTACTCGAGACTTTTGTACAGTGAAGTCTCTCGAGCCCTGCTCTCTACTT 5431
 QY 682 GATCGATACGCGGTTGAGCATGTTAAATTTGGAAGTGTGTCAGAGTGAATGATTAAGA 741
 DB 5430 AAGCATTTATCCGAAACGCAACACATTAACATAGGCTCCGGGGGGAATAAGTATCAT 5371
 QY 742 GAGTTGGCTGAGTTGTGTAAGGTTGTGTTTGAAGGAGAGCTTGGATGGGATTC 801
 DB 5370 GAACTAGCCCATGCTGCGCGTGTGTTGTTTAAAGGCGATATAGTCTTCGACACA 5311
 QY 802 ACTAAGCCAGATGGCACACCGAGGAACCTTATGGACAGCTCAAGCTCGCGTCTTGGT 861
 DB 5310 TCCNAGCCGACGGAACCGCACGAAGCTTTTATCTAGCGAAGACTCGTGTGATGGT 5251
 QY 862 TGAACACCTAAGGTTTCTCTAGAGATGCTGAGCCAACTTATGATTTGATTTGAAG 921
 DB 5250 TGGCGCCGCAAGACCTCGCTCGAGCTGGGACTGGCCAAATCTCTATGAATCGTTGTG 5191
 QY 922 AATGT 926
 DB 5190 AATGT 5186

RESULT 15
 ID AAV30459/C
 AC AAV30459 standard; DNA; 536165 BP.
 XX
 AC AAV30459;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Rhizobium species symbiotic plasmid pNGR234.
 XX
 KW Symbiosis; open reading frame; ORF; plasmid; vector; transportation;

KW degradation; metabolism; host range; nitrogen fixation; nodulation;
 KW legume; plant; ds.
 XX Rhizobium sp.
 OS
 XX WO9802560-A2.
 XX PD 22-JAN-1998.
 XX PF 10-JUL-1997; 97WO-1B00950.
 XX PR 20-MAY-1997; 97GB-0010395.
 PR 12-JUL-1996; 96EP-0730001.
 XX (MOLE-) INST MOLECULAR BIOTECHNOLOGY.
 PA (BIOL-) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.
 XX Broughton WJ, Freiberg CB, Perret XP, Rosenthal A;
 PI WPI; 1998-110606/10.
 DR
 XX
 PT New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to
 PT develop products for modifying plant characteristics, e.g. nitrogen
 PT fixation, synthesis of compounds and stress response
 XX
 PS Claim 1; Fig 3; 228pp; English.
 XX
 CC This is the nucleotide sequence of the plasmid pNGR234a isolated from
 CC Rhizobium sp. NGR234. Open reading frames (ORF) derivable from the
 CC nucleotide sequence are claimed. The nucleotide sequences or ORFs can
 CC be used e.g. in the transportation of compounds to and from an organism
 CC which is a host to at least one of the nucleotide sequences, ORFs or
 CC proteins, the degradation and/or metabolism of organic, inorganic,
 CC natural or xenobiotic substances in a host organism or the modification
 CC of the host range, nitrogen fixation abilities; for obtaining a synthetic
 CC minimal set of ORFs required for functional Rhizobium-legume symbiosis,
 CC especially for nodulation efficiency on host plants.
 XX
 SQ Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;

Query Match 28.3%; Score 265; DB 19; Length 536165;
 Best Local Similarity 56.6%; Pred. No. 3e-73;
 Matches 512; Conservative 0; Mismatches 390; Indels 3; Gaps 1;
 QY 22 ATCTTCGTCGGGTTCATCGTGTGTTGGTTCGATCTGCAATTTGTCGCAAGCTTCAGGAA 81
 DB 6087 ATTTGGTTCGAGGACACAAAGGTATGTCGCGAGCCATTAATTCGATCGTTCCTCC 6028
 QY 82 CAAGGTTTACCAATCTCGTCTTAAACACACGCGAGCTTGATCTCACTCGTCAAGCC 141
 DB 6027 GAGGATTGCGAAGTCATCGTTGCAGATAGCA---AAAGCTTGATCTGACGCGCAAGAG 5971
 QY 142 GATGTTGAATCTCTTTTCTCAAGAGAACGAGTTTATGTAATCCTAGCAGCAGCTAAA 201
 DB 5970 GAAGTTGAGAAATTTCTATTAAAGGAAAGCCGACGCGGTCTAATATGCGCGCAGGAAG 5911
 QY 202 GTTGGTGTATTACGCTTAACACACCTATCTCTGCTGATTTTCATTTGTTGTCAATCTCCAG 261
 DB 5910 GTGGTGGGATCTTGGCAATGATATGATGCGCGTCTATGCTATCTATCAAAAACCTTATC 5851
 QY 262 ATTACAGACCAATGTGATCCACTCTGCAATATGAGCACGCGTGTGAAGAGCTTCTCTTCT 321
 DB 5850 ATGGAGGCTAATGTCATTGAGGGCTCTCTTCGCGAGTGGCGTTGAAAAGCTTCTTTTCT 5791
 QY 322 GGATCATCTGCAATTTACCTTAAATTTGCTCTCTAGCCAAATTCCTGAGTCTGCTTTGTTA 381
 DB 5790 GGATCGAGTTGCATATATCCGAAGTATGCGCGCAGCCCAATAAGGAAGAGGCTCTATTA 5731
 QY 382 ACAGATCTGCTGAACCAACTAATGAGTGTGATGCTATTGCTTAAGATCGCTGGGATTAAG 441
 DB 5730 ACCGACCACTTGAGCGGACCAACGAGTGTGATGCGATCGCCAAATCGCGGCATTAA 5671
 QY 442 ACTTGTACGCTTATAGGATTCAGACGCGATGGGATGCAATCTCTGCGCATGCTCTACTAAT 501

```

Db 5670 TTGTGTCAAGCGTATCGTAAGCAATACGGCGAACTTCATATCAGCCATGCCGAAAT 5611
QY 502 CTCTATGTCCTAATGACAAATTTCCACCGGAGAAATCTCATGTGCTTCTGCTCTTATG 561
Db 5610 CTCTATGGCCACGGCATTAAGTTGATCTTAACCTCAGCCACGTGCTCCCTGCCTTAATA 5551
QY 562 AGGAGTTTCCAGAGCGGAAAGTGAATGAGCGGAGAAAGTTGTGTTGGGGTACAGGT 621
Db 5550 CGCAAGACACATGAGGCAAGATTTAAGACCTTGGGTGCTTGTCTATATGGGGAAGCGGC 5491
QY 622 AGTCGGTTGAGGAGTTCTTGCAATGTTGATGATTTGGCTGATGCTTGTGTTTCTTCTGCTG 681
Db 5490 ACACCTACTCGAGACTTTTGTACAGTGAAGACTGCTCCGAGCCCTGGTCTTCTTACTT 5431
QY 682 GATCGATACAGCGGTTGGAGCATGTTAAACATTTGGAAGTGGTCAAGAAGTGACTATTAGA 741
Db 5430 AAGCATTTATCCGAAACGGACACATTAACATAGGCTCCGGGGGGGAATAGTATCATT 5371
QY 742 GAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAGGGAAGCTTCGATGGGATTC 801
Db 5370 GAACTAGCCACATCGTCTCGCGTGTGTTGGTTTAAAGGCGATATAGTCTTCGACACA 5311
QY 802 ACTAGCCAGATGGCACACGGAGAACTTATCGACAGCTCAAAGCTCGGCTCTTTGGGT 861
Db 5310 TCCAGCCGAGCGGAACGCCACGAAAGCTTTTATAGCGAAAGACTCGTGTGATGGGT 5251
QY 862 TGGACACCTAAGGTTTCTCTTAGAGATGCTCTGAGCCAAACTTATGATGGTATTTGAAG 921
Db 5250 TGGGCCCCGAGACCTCGCTCGAGCTGGGACTGGCCCAATCCTATGAATCGTTTGTACGC 5191
QY 922 AATGT 926
Db 5190 AATGT 5186

```

Search completed: June 3, 2003, 02:47:57
Job time : 300 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 3, 2003, 02:38:25 ; Search time 2637 Seconds
(without alignments)
10330.006 Million cell updates/sec

Title: US-10-089-014-2
Perfect score: 936
Sequence: 1 atgtgtgacaaatctgccaa.....tgagaagtgttgcacacga 936

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rnd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	936	100.0	939	8	AB034806	AB034806 Arabidops
2	936	100.0	939	8	AF045286	AF045286 Arabidops
3	936	100.0	70836	8	AC010556	AC010556 Arabidops
4	667.8	71.3	1018	8	AY096364	AY096364 Arabidops
5	667.8	71.3	1341	8	AY063880	AY063880 Arabidops
6	667.8	71.3	65262	2	AC025908	AC025908 Arabidops
7	667.8	71.3	75133	8	F2H15	AC034106 Sequence
8	659	70.4	1353	8	AY089031	AY089031 Arabidops
9	430.6	46.0	130393	2	AP003565	AP003565 Oryza sat
10	391	41.8	1490	6	AX172657	AX172657 Sequence
11	354	37.8	10029	1	AE013290	AE013290 Methanosa
12	301.8	32.2	8458	1	AE007718	AE007718 Clostridi
13	297	31.7	1020	6	A63794	A63794 Sequence 15
14	297	31.7	4435	1	MAP223833	AJ223833 Mycobacte
15	297	31.7	4435	6	A63783	A63783 Sequence 4
16	295.4	31.6	1020	6	A63792	A63792 Sequence 13
17	295.4	31.6	7995	6	A63782	A63782 Sequence 3
18	295.4	31.6	8938	1	MAS223832	AJ223832 Mycobacte
19	295.4	31.6	25861	1	AF125999	AF125999 Mycobacte
20	295.4	31.6	71286	1	AF143772	AF143772 Mycobacte
21	285.4	30.5	130393	2	AP003565	AP003565 Oryza sat
22	284.6	30.4	10999	1	AE013711	AE013711 Yersinia
23	284.6	30.4	20588	1	YPS251712	AJ251712 Yersinia
24	279.4	29.9	26913	1	YEU46859	U46859 Yersinia en
25	276.8	29.6	19553	1	AE007023	AE007023 Mycobacte
26	276.4	29.5	330743	14	PRU42580	U42580 Paramecium
27	275.2	29.4	12732	6	AR128139	AR128139 Sequence
28	275.2	29.4	37316	1	MSGY456	AD000001 Mycobacte
29	275.2	29.4	38300	1	MTIC277	Z79701 Mycobacteri
30	274.2	29.3	20745	1	YPE251713	AJ251713 Yersinia
31	274.2	29.3	205050	1	AJ414155	AJ414155 Yersinia
32	271.8	29.0	20148	1	AE008793	AE008793 Salmonell
33	271.8	29.0	250050	1	AL627273	AL627273 Salmonell
34	270.2	28.9	12016	1	AE009407	AE009407 Agrobacte
35	270.2	28.9	12862	1	AE008206	AE008206 Agrobacte
36	269.8	28.8	12666	1	AE005429	AE005429 Escherich
37	269.8	28.8	14002	1	AF061251	AF061251 Escherici
38	269.8	28.8	14029	6	AX029262	AX029262 Sequence
39	269.8	28.8	14187	6	AR204225	AR204225 Sequence
40	269.8	28.8	31482	1	AB008676	AB008676 Escherich
41	269.8	28.8	286857	1	AP002559	AP002559 Escherich
42	268.6	28.7	12897	1	AF285084S1	AF285084 Salmonell
43	265	28.3	11429	1	AE000064	AE000064 Rhizobium
44	265	28.3	320040	6	A79350	A79350 Sequence 1
45	265	28.3	320040	6	A93002	A93002 Sequence 1

ALIGNMENTS

RESULT 1
AB034806
LOCUS
DEFINITION
AB034806
Arabidopsis thaliana AtFX mRNA for
GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase, complete
cds.
ACCESSION
AB034806
VERSION
AB034806.1 GI:7804473
KEYWORDS
GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase.
SOURCE
Arabidopsis thaliana
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE
1 (sites)

AUTHORS Nakayama, K., Maeda, Y., Wang, X. and Jigami, Y.
 TITLE Expression of the genes for GDP-fucose synthesis in yeast
 JOURNAL Saccharomyces cerevisiae
 REFERENCE Unpublished
 AUTHORS 2 (bases 1 to 939)
 TITLE Nakayama, K. and Jigami, Y.
 JOURNAL Direct Submission
 Submitted (09-NOV-1999) Ken-ichi Nakayama, National Institute of
 Bioscience and Human Technology, Molecular Biology Department, 1-1
 Higashi, Tsukuba, Ibaraki 305-8566, Japan
 (E-mail: nakayama@nibh.go.jp, Tel: 81-298-54-6226,
 Fax: 81-298-54-6220)

FEATURES
 source Location/Qualifiers
 1..939
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 1..939
 /gene="AtFX"
 1..939
 /gene="AtFX"
 /codon_start=1
 /product="GDP-4-keto-6-deoxy-D-mannose-3,
 5-epimerase-4-reductase"
 /protein_id="BAA95670.1"
 /db_xref="GI:780474"
 /translation="MSDKSAKIFVAGHRLVGSIAVRKLEQSGTNNLVKTHAELDLT
 ROADVSFQSEKPVVILAAKVGVIHANNTVPADFIGVNIQIOTNLVHSAVEHGVK
 KLLFLGSSCIYKPKFAPQIPESALLTASLEPNEWYAIKAGIKTCQAYRLOHGDA
 ISGMPTLVGPNDFPNSHVLPAUMRFHAKVNGAEVVMVGTSPLRFLHVDV
 LADACVFLDRYDGLSHVNIQSGQEVITRELAEVLKEVVGFEKGLGWDCTKPDGTPRK
 LMDSSKSLASLGWTPKVSRLDGLSGQTDYDLKNCNR"
 BASE COUNT 229 a 190 c 240 g 280 t

Query Match 100.0%; Score 936; DB 8; Length 939;
 Best Local Similarity 100.0%; Pred. No. 1.2e-265;
 Matches 936; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTTGGATCTGCC 60
 DB 1 ATGCTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTTGGTTGGATCTGCC 60
 QY 61 ATTGTCGGCAAGCTTCAGGAACAAGTTTCCACCAATCTCGTTCTTAAACACACACCGCGAG 120
 DB 61 ATTGTCGGCAAGCTTCAGGAACAAGTTTCCACCAATCTCGTTCTTAAACACACACCGCGAG 120
 QY 121 CTTGATCTCACTGTCAGCCGATGTGATCCTCTTTTCAAGAGAAGCCAGTTTAT 180
 DB 121 CTTGATCTCACTGTCAGCCGATGTGATCCTCTTTTCAAGAGAAGCCAGTTTAT 180
 QY 181 GTAACTCTAGCAGCAGTAAAGTTGGTGGTATTTCAGCTAAACACACCTATCTGCTGAT 240
 DB 181 GTAACTCTAGCAGCAGTAAAGTTGGTGGTATTTCAGCTAAACACACCTATCTGCTGAT 240
 QY 241 TTCAATGGTGTCAATCTCAGATTCAGACCAATGTGATCCACTCTGCATATGACGACGCT 300
 DB 241 TTCAATGGTGTCAATCTCAGATTCAGACCAATGTGATCCACTCTGCATATGACGACGCT 300
 QY 301 GTGAAGAAGCTTCTCTCTTGGATCATCTGATTAACCTTAAATTTGCTCTCAGCCA 360
 DB 301 GTGAAGAAGCTTCTCTCTTGGATCATCTGATTAACCTTAAATTTGCTCTCAGCCA 360
 QY 361 ATTCCTGAGTCTCTTGTGTTTAAACAGCATCGTTTGAACCAACTAAATGAGTGGTATGATT 420
 DB 361 ATTCCTGAGTCTCTTGTGTTTAAACAGCATCGTTTGAACCAACTAAATGAGTGGTATGATT 420
 QY 421 GCTAAGATCGTGGGATTAAGACTTCTCAGGCTTATAGGATTCAGACCGATGGGATGCA 480
 DB 421 GCTAAGATCGTGGGATTAAGACTTCTCAGGCTTATAGGATTCAGACCGATGGGATGCA 480
 QY 481 ATCTCTGGCATGCTACTAATCTCTATGTCCTTAATGACAATTTCCACCGGAGATTCT 540
 DB 481 ATCTCTGGCATGCTACTAATCTCTATGTCCTTAATGACAATTTCCACCGGAGATTCT 540

QY 541 CATGTGCTTCCTGCTCTTATGAGGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGAA 600
 DB 541 CATGTGCTTCCTGCTCTTATGAGGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGAA 600
 QY 601 GTTGTGGTGTGGGTACAGGTAGTCCGTTGAGGGAGTCTTTCATGTTGATGATTGGCT 660
 DB 601 GTTGTGGTGTGGGTACAGGTAGTCCGTTGAGGGAGTCTTTCATGTTGATGATTGGCT 660
 QY 661 GATGCTTGTTTCTTCTGCTGATCGATACAGCGGGTTGGAGCATGTTAAACATTGGAAGT 720
 DB 661 GATGCTTGTTTCTTCTGCTGATCGATACAGCGGGTTGGAGCATGTTAAACATTGGAAGT 720
 QY 721 GGTCAAGAAGTCACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGAGTCTTGGTTTGA 780
 DB 721 GGTCAAGAAGTCACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGAGTCTTGGTTTGA 780
 QY 781 GGGAAAGCTTGGATGGATTGCACTAAGCCAGATGCGACACCGAGGAAACTTATGGACAGC 840
 DB 781 GGGAAAGCTTGGATGGATTGCACTAAGCCAGATGCGACACCGAGGAAACTTATGGACAGC 840
 QY 841 TCAAAAGCTCGGCTCTTGGGTTGGACACCTAAGGTTCTCTTAGAGATGGTCTGAGCCAA 900
 DB 841 TCAAAAGCTCGGCTCTTGGGTTGGACACCTAAGGTTCTCTTAGAGATGGTCTGAGCCAA 900
 QY 901 ACTTATGATTGGTATTGGAAGAAATGTTTGCACACCGA 936
 DB 901 ACTTATGATTGGTATTGGAAGAAATGTTTGCACACCGA 936
 RESULT 2
 AF045286 939 bp DNA linear PLN 19-APR-2000
 LOCUS Arabidopsis thaliana
 DEFINITION GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase (GER1) gene,
 complete cds.
 ACCESSION AF045286
 VERSION AF045286.3 GI:7596994
 KEYWORDS Arabidopsis thaliana.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 939)
 AUTHORS Bonin, C.P. and Reiter, W.D.
 TITLE A bifunctional epimerase-reductase acts downstream of the MUR1 gene
 product and completes the de novo synthesis of GDP-L-fucose in
 Arabidopsis
 JOURNAL Plant J. 21 (5), 445-454 (2000)
 MEDLINE 20223138
 PUBMED 10758496
 REFERENCE 2 (bases 1 to 936)
 AUTHORS Bonin, C.P., Potter, I., Vanzin, G.F. and Reiter, W.-D.
 TITLE Direct Submission
 JOURNAL Submitted (29-JAN-1998) Molecular and Cell Biology, University of
 Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA
 REFERENCE 3 (bases 1 to 936)
 AUTHORS Bonin, C.P.
 TITLE Direct Submission
 JOURNAL Submitted (08-OCT-1999) Molecular and Cell Biology, University of
 Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA
 REFERENCE 4 (bases 1 to 939)
 AUTHORS Reiter, W.-D.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-2000) Molecular and Cell Biology, University of
 Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA
 REFERENCE 5 (bases 1 to 939)
 AUTHORS Reiter, W.-D.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-2000) Molecular and Cell Biology, University of
 Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA
 REMARK On Apr 19, 2000 this sequence version replaced gi:6016478.
 COMMENT Location/Qualifiers
 1..939
 /organism="Arabidopsis thaliana"

Gene	/cultivar="Columbia ecotype"	
	/db_xref="taxon:3702"	
mRNA	<1..>939	
	/gene="GER1"	
CDS	<1..>939	
	/gene="GER1"	
CDS	/product="GDP-4-keto-6-deoxy-D-mannose-3, 5-epimerase-4-reductase"	
	1..939	
CDS	/note="enzyme in pathway for de novo synthesis of GDP-L-fucose"	
	/codon_start=1	
CDS	/product="GDP-4-keto-6-deoxy-D-mannose-3, 5-epimerase-4-reductase"	
	/protein_id="AAC02703.2"	
CDS	/db_xref="GI:6016479"	
	/translation="MSDSKAKIFVAGHGLVSAIVKLOEGFTNLVLTAEIDLIT	
CDS	ROADVESFRSQKPVYVILAAKVGCIHANNTYPAFIGVNLQIQTNVHSAIEHGVK	
	KLLFGLSSCIYKPKFAPQIPESALLTASLEPTNEWYAIAGIKTCQYRIQHGWD	
CDS	ISGMPTNLYPNDRHPENSHVLPALMRREHAKVNGAEVYVWGTGSPLEFLHVD	
	LADACVFLIDRYSLGHEVNIIGSQEVTIRELAELVKEVVFEGKLGWDCTKPDGTPRK	
CDS	LMDSSKLSLGTWPKVSLRDGLSOTVDWYLNKVCNR"	
	229 a 190 c 240 g 280 t	
BASE COUNT	Query Match	
	Best Local Similarity 100.0%; Score 936; DB 8; Length 939;	
ORIGIN	Matches 936; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
	1 ATGCTGACAAATCTCCAAAATCTTCGTCGGGGTCATCGTGGTTGGTTGATCTGCC 60	
QY	1 ATGCTGACAAATCTCCAAAATCTTCGTCGGGGTCATCGTGGTTGGTTGATCTGCC 60	
	61 ATTGTCGCGAAGCTTCAGGAACAAGTTTCCACAACTCTGTTCTTAAACACACGCCGAG 120	
DB	61 ATTGTCGCGAAGCTTCAGGAACAAGTTTCCACAACTCTGTTCTTAAACACACGCCGAG 120	
	121 CTTGATCTCACTCGTCAAGCCGATGTTGAATCTCTTTTCAAGAGAAGCAGTTTAT 180	
QY	121 CTTGATCTCACTCGTCAAGCCGATGTTGAATCTCTTTTCAAGAGAAGCAGTTTAT 180	
	181 GTAATCTAGCAGCACTAAAGTTGGTGTATTCAGCTTAAACACCACTATCTGCTGAT 240	
DB	181 GTAATCTAGCAGCACTAAAGTTGGTGTATTCAGCTTAAACACCACTATCTGCTGAT 240	
	241 TTCATTGGTGTCAATCTCCAGATTTCAGCAATATGTCATCTCCATATGAGCAGGT 300	
QY	241 TTCATTGGTGTCAATCTCCAGATTTCAGCAATATGTCATCTCCATATGAGCAGGT 300	
	301 GTGAAGAAGCTTCTCTTCTTGGATCATCTGCAATTAACCTAAATTTGCTCTCAGCCA 360	
DB	301 GTGAAGAAGCTTCTCTTCTTGGATCATCTGCAATTAACCTAAATTTGCTCTCAGCCA 360	
	361 ATTCCTGAGTCTGCTTTGTTAAACAGATCGTTGAACCACTAATGAGTGGTATGCTATT 420	
QY	361 ATTCCTGAGTCTGCTTTGTTAAACAGATCGTTGAACCACTAATGAGTGGTATGCTATT 420	
	421 GCTAAGATCGTGGGATTAAGCTTCTCAGGCTTATAGGATTCAGACCGGATGGATGCA 480	
DB	421 GCTAAGATCGTGGGATTAAGCTTCTCAGGCTTATAGGATTCAGACCGGATGGATGCA 480	
	481 ATCTCTGGCATGCTACTAATCTCTATGGTCTTAATGACAATTTCCACCGGAGATTCT 540	
QY	481 ATCTCTGGCATGCTACTAATCTCTATGGTCTTAATGACAATTTCCACCGGAGATTCT 540	
	541 CATGTGCTTCTCTCTTATGAGGATTCACAGCGCGAAAGTGAATGAGCGGAGGAA 600	
DB	541 CATGTGCTTCTCTCTTATGAGGATTCACAGCGCGAAAGTGAATGAGCGGAGGAA 600	
	601 GTTGTGGTGGGGTACAGGTAGTCCGTTGAGGAGTCTTTCATGATGATTTGGCT 660	
QY	601 GTTGTGGTGGGGTACAGGTAGTCCGTTGAGGAGTCTTTCATGATGATTTGGCT 660	
	660 GTTGTGGTGGGGTACAGGTAGTCCGTTGAGGAGTCTTTCATGATGATTTGGCT 660	

QY	661	GATGCTTGTTGTTTCTGCTGATCGATACAGCGGTTGGACATGTTAAACATTGGAAGT	720						
Db	661	GATGCTTGTTGTTTCTGCTGATCGATACAGCGGTTGGACATGTTAAACATTGGAAGT	720						
QY	721	GCTCAAGAAGTGACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGA	780						
Db	721	GCTCAAGAAGTGACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGA	780						
QY	781	GGGAAGCTTGGATGGATTGCACTAAGCCAGATGGCACACCGAGGAACTTATGGACAGC	840						
Db	781	GGGAAGCTTGGATGGATTGCACTAAGCCAGATGGCACACCGAGGAACTTATGGACAGC	840						
QY	841	TCAAAGCTCGCGTCTTTGGTTGGACACCTAAGGTTTCTCTTAGAGATGGTCTGAGCCAA	900						
Db	841	TCAAAGCTCGCGTCTTTGGTTGGACACCTAAGGTTTCTCTTAGAGATGGTCTGAGCCAA	900						
QY	901	ACTTATGATTTGGTATTTCGAAGATGTTTGGCAACCGA	936						
Db	901	ACTTATGATTTGGTATTTCGAAGATGTTTGGCAACCGA	936						
RESULT 3	AC010556	70836 bp	DNA	linear	PLN 15-JUN-2001				
LOCUS	AC010556	Arabidopsis thaliana chromosome 1 BAC T18K17 genomic sequence,							
DEFINITION	AC010556	complete sequence.							
ACCESSION	AC010556	GI:12324309							
VERSION	HTG								
KEYWORDS	Arabidopsis thaliana.								
SOURCE	Arabidopsis thaliana								
ORGANISM	Arabidopsis thaliana								
REFERENCE	1 (bases 1 to 70836)	Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J., Wu,D., Matti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M. Arabidopsis thaliana chromosome 1 BAC T18K17 genomic sequence							
AUTHORS	Unpublished								
TITLE	2 (bases 1 to 70836)	Lin,X. and Kaul,S.							
REFERENCE	Direct Submission								
AUTHORS	Submitted (15-SEP-1999)	The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org							
JOURNAL	3 (bases 1 to 70836)	Town,C.D. and Kaul,S.							
REFERENCE	Direct Submission								
AUTHORS	Submitted (19-JAN-2001)	The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org							
JOURNAL	On Jan 19, 2001	this sequence version replaced gi:12280860. Address all correspondence to:at@tigr.org							
COMMENT	BAC clone T18K17 is from Arabidopsis thaliana chromosome 1								
The orientation of the sequence is from SP6 to T7 end of the BAC clone.									
Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky, http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, http://www.tigr.org/softlab/glimmerM.html, and GenesSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as									

hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

source	Location/Qualifiers
	1. 70836
	/organism="Arabidopsis thaliana"
	/cultivar="Columbia"
	/db_xref="taxon:3702"
	/chromosome="1"
	/clone="T18K17"
repeat_region	complement (382..402)
gene	/rpt_family="AT_rich"
	516..1979
	/gene="T18K17.1"
	/note="predicted by gensecan"
mRNA	join(<516..836,1121..1263,1342..1526,1807..>1979)
CDS	/gene="T18K17.1"
	join(516..836,1121..1263,1342..1526,1807..1979)
	/gene="T18K17.1"
	/codon_start=1
	/product="hypothetical protein; 516-1979"
	/protein_id="AAG52120.1"
	/db_xref="GI:12324311"
	/translation="MEMEAAEIVHSAARFEAEKMGKEDAVEEIRMSGYGGDIVVWG GPSLNTPTGVTGSMVMDSGVLGKPLEHSGVDSKVLKLGKIVELSGGGLVGLTAAL SPCSLNTPTGVTGSMVMDSGVLGKPLEHSGVDSKVLKLGKIVELSGGGLVGLTAAL LGNNAVLTDLPELRLLKNIQTNLHRCNTRGSATVQELVWGGDDPDLIERPPDYDA VLEVFETALKDFAIGRVEQTQHPDYRSHRVVLVLEKKRCLADESSLNQSC"
gene	2530..4892
	/gene="T18K17.2"
	/note="similar to serine carboxypeptidase I precursor GB:P07519 [Hordeum vulgare], glucose acyltransferase GB:AD01263 [Solanum berthaultii]; contains Pfam profile: PF00450 Serine carboxypeptidase"
mRNA	join(<2530..2678,2770..2915,2998..3074,3151..3273, 3357..3459,3546..3685,3762..3848,3946..4014,4096..4167, 4237..4355,4441..4554,4650..4695,4812..>4892)
	/gene="T18K17.2"
CDS	join(2530..2678,2770..2915,2998..3074,3151..3273, 3357..3459,3546..3685,3762..3848,3946..4014,4096..4167, 4237..4355,4441..4554,4650..4695,4812..>4892)
	/gene="T18K17.2"
	/codon_start=1
	/product="putative serine carboxypeptidase; 2530-4892"
	/protein_id="AAG52126.1"
	/db_xref="GI:12324317"
	/translation="MANNVSVLVKSLLLHLVFLSKQHVDSASIVKPLPGFEGPLP FELETYIGVEEVEVQVLYFIKSRNPKEDPLLLWLTGGPGCSAISGLLYQNGPLA MKLDVNGTLPISLVTSYTWKTSMTFLDQPVGTGFSYRTQFNKPSDGEAKRIH EFLQKLGKHQFSSNPFYVAGDSYGLVVPATVQVEISKNGCCNRPINLQGYVLGN PLTDCVDCNVRVPAHGMALISDELVESLKCTCKGEYVNVHPDTECLKFVEEFNKL TRVCRHILHSCCETETSCSYRFLTYTWANDTVKALQINKESIGEWRCYCYRG IPYNDIKSSPMYHNNISIDYRSILYSGDHLNVPFLATQAWRLSNYSIIDNWRPWWIK MIKQIAGTYTRYANKMTFATVGGGHTAEFTPEKTFMFWINGQPL"
repeat_region	complement (4703..4758)
gene	/rpt_family="(A)n"
	5659..8034
	/gene="T18K17.3"
	/note="similar to serine carboxypeptidase I precursor GB:P07519 [Hordeum vulgare], glucose acyltransferase GB:AD01263 [Solanum berthaultii]; contains Pfam profile: PF00450 Serine carboxypeptidase; non-consensus donor splice site GA at exon 8"
mRNA	join(<5659..5807,5899..6044,6153..6229,6315..6437, 6527..6629,6714..6853,6937..7023,7122..7190,7278..7349, 7429..7547,7619..7732,7814..7859,7954..>8034)
	/gene="T18K17.3"
	/note="contains non-consensus splice site"
CDS	join(5659..5807,5899..6044,6153..6229,6315..6437, 6527..6629,6714..6853,6937..7023,7122..7190,7278..7349, 7429..7547,7619..7732,7814..7859,7954..>8034)
	/gene="T18K17.3"
	/codon_start=1
	/product="putative serine carboxypeptidase; 12385-14737"
	/protein_id="AAG52138.1"
	/db_xref="GI:12324329"
	/translation="MANNVSVLVKSLLLHLVFLSKQHVDSASIVKPLPGFEGPLP FELETYIGVEEVEVQVLYFIKSRNPKEDPLLLWLTGGPGCSAISGLLYQNGPLA MKLDVNGTLPISLVTSYTWKTSMTFLDQPVGTGFSYRTQFNKPSDGEAKRIH EFLQKLGKHQFSSNPFYVAGDSYGLVVPATVQVEISKNGCCNRPINLQGYVLGN PLTDCVDCNVRVPAHGMALISDELVESLKCTCKGEYVNVHPDTECLKFVEEFNKL TRVCRHILHSCCETETSCSYRFLTYTWANDTVKALQINKESIGEWRCYCYRG IPYNDIKSSPMYHNNISIDYRSILYSGDHLNVPFLATQAWRLSNYSIIDNWRPWWIK MIKQIAGTYTRYANKMTFATVGGGHTAEFTPEKTFMFWINGQPL"
repeat_region	complement (10483..10509)
gene	/rpt_family="(TA)n"
	11537..11568
	/rpt_family="(A)n"
	/complement(11583..11712)
repeat_region	/rpt_family="AT_rich"
	complement(11785..11807)
repeat_region	/rpt_family="AT_rich"
	12385..14737
	/gene="T18K17.5"
	/note="similar to serine carboxypeptidase I precursor GB:P07519 [Hordeum vulgare], glucose acyltransferase GB:AD01263 [Solanum berthaultii]; contains Pfam profile: PF00450 Serine carboxypeptidase"
mRNA	join(<12385..12533,12631..12776,12858..12934,12999..13121, 13208..13310,13426..13565,13643..13729,13820..13888, 13969..14040,14131..14249,14327..14440,14520..14565, 14657..>14737)
	/gene="T18K17.5"
	join(12385..12533,12631..12776,12858..12934,12999..13121, 13208..13310,13426..13565,13643..13729,13820..13888, 13969..14040,14131..14249,14327..14440,14520..14565, 14657..>14737)
	/gene="T18K17.5"
	/codon_start=1
	/product="putative serine carboxypeptidase; 12385-14737"
	/protein_id="AAG52138.1"
	/db_xref="GI:12324329"
	/translation="MANNVSVLVKSLLLHLVFLSKQHVDSASIVKPLPGFEGPLP FELETYIGVEEVEVQVLYFIKSRNPKEDPLLLWLTGGPGCSAISGLLYQNGPLA MKLDVNGTLPISLVTSYTWKTSMTFLDQPVGTGFSYRTQFNKPSDGEAKRIH EFLQKLGKHQFSSNPFYVAGDSYGLVVPATVQVEISKNGCCNRPINLQGYVLGN PLTDCVDCNVRVPAHGMALISDELVESLKCTCKGEYVNVHPDTECLKFVEEFNKL TRVCRHILHSCCETETSCSYRFLTYTWANDTVKALQINKESIGEWRCYCYRG IPYNDIKSSPMYHNNISIDYRSILYSGDHLNVPFLATQAWRLSNYSIIDNWRPWWIK MIKQIAGTYTRYANKMTFATVGGGHTAEFTPEKTFMFWINGQPL"

MKLDVNGTLPVSLVSTTYSWTYKASSMIFLDQPVGAGFSYRSRTQLLNKPSDSGEAKRIH BFLQWLGHQBFSSNPFYVGDSDGSMVVPATVQBIISKGNVECCNPPINLQGYLGN PLTFDYDYNLSIFPAHGMALISDELFSKIDKCTKGYRNVHRNTECLKLFIEBENK TNSICORRIIDPFCTPCNVIYRFLAAAYWANDETVRKALQIKKETIGEWVRCHYG IPYNDIKSSIPYHMNNSINGVRSIYSDGHDHFEVFLGTQAWIRSLNYSVIDDNRPW									
Query Match	100.0%	Score	936;	DB	8;	Length	70836;		
Best Local Similarity	100.0%	Pred. No.	2.1e-265;						
Matches	936;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	ATGTCGTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTGTGGTTGGTGGATCTGCC	60						
DB	21556	ATGTCGTGACAAATCTGCCAAATCTTCGTCGGGGTCATCGTGGTGTGGTTGGTGGATCTGCC	21615						
QY	61	ATTGTCGCGAAGTTTCAGGAACAAGTTTTCACCAATCTCGTTCTTTAAACACACGCGCGAG	120						
DB	21616	ATTGTCGCGAAGTTTCAGGAACAAGTTTTCACCAATCTCGTTCTTTAAACACACGCGCGAG	21675						
QY	121	CTTGATCTCACTCGTCAAGCCGATGTTGAATCCTTCTTCTTCAAGAGAAGCCAGTTTAT	180						
DB	21676	CTTGATCTCACTCGTCAAGCCGATGTTGAATCCTTCTTCTTCAAGAGAAGCCAGTTTAT	21735						
QY	181	GTAAATCCTAGCAGCAGTAAAGTTGGTGGTATTTCAGCTAAACAACCTATCTCTGCTGAT	240						
DB	21736	GTAAATCCTAGCAGCAGTAAAGTTGGTGGTATTTCAGCTAAACAACCTATCTCTGCTGAT	21795						
QY	241	TTCATTGGTCAATCTCAGATTTCAGACCAATGTGATCCACTCTGCATATGAGCAGCGT	300						
DB	21796	TTCATTGGTCAATCTCAGATTTCAGACCAATGTGATCCACTCTGCATATGAGCAGCGT	21855						
QY	301	GTGAAGAAGCTTCTTCTTGGATCATCTGATTTACCTTAATCCCTAAATTTGCTCTCAGCCA	360						
DB	21856	GTGAAGAAGCTTCTTCTTGGATCATCTGATTTACCTTAATCCCTAAATTTGCTCTCAGCCA	21915						
QY	361	ATTCTCGAGTCTGCTTGTGTTTAAACAGCATCGTCTTGACCAACTAAATGAGTGGTATGTA	420						
DB	21916	ATTCTCGAGTCTGCTTGTGTTTAAACAGCATCGTCTTGACCAACTAAATGAGTGGTATGTA	21975						
QY	421	GCTAAGATCGCTGGGATTAAGACTTGTTCAGGCTTATAGGATTCAGACGGATGGGATGCA	480						
DB	21976	GCTAAGATCGCTGGGATTAAGACTTGTTCAGGCTTATAGGATTCAGACGGATGGGATGCA	22035						
QY	481	ATCTCTGGATGCTACTAATCTCTATGCTCTATGATGACAAATTTCCACCCGGAGATTCT	540						
DB	22036	ATCTCTGGATGCTACTAATCTCTATGCTCTATGATGACAAATTTCCACCCGGAGATTCT	22095						
QY	541	CATGCTCTTCTGCTCTTATGAGGAGTTCCAGGCGCAAGTGAATGAGCGGAGGAA	600						
DB	22096	CATGCTCTTCTGCTCTTATGAGGAGTTCCAGGCGCAAGTGAATGAGCGGAGGAA	22155						
QY	601	GTGTGGTGTGGGTCACAGTATGCTGCTGAGGAGTTCTTGCATGTTGATGTTGGCT	660						
DB	22156	GTGTGGTGTGGGTCACAGTATGCTGCTGAGGAGTTCTTGCATGTTGATGTTGGCT	22215						
QY	661	GATGCTTGTGTTTCTTGTGATCATGATCAGCGGTTGGAGATGTTAACTTGAAT	720						
DB	22216	GATGCTTGTGTTTCTTGTGATCATGATCAGCGGTTGGAGATGTTAACTTGAAT	22275						
QY	721	GGTCAAGAAGTACTATTAGAGTTGGTGGTGGTCAAGAGAGTTGTTGGTTTGA	780						
DB	22276	GGTCAAGAAGTACTATTAGAGTTGGTGGTGGTCAAGAGAGTTGTTGGTTTGA	22335						
QY	781	GGGAAGCTTTGGATGGATGCTACTAAGCCAGATGGCACACCGAGGAACTTATGGACAGC	840						
DB	22336	GGGAAGCTTTGGATGGATGCTACTAAGCCAGATGGCACACCGAGGAACTTATGGACAGC	22395						
QY	841	TCAAAGCTCGGCTTTTGGTGGTGGACCACTAAGTTTCTTCTTAGAGATGCTCTGAGCCAA	900						
DB	22396	TCAAAGCTCGGCTTTTGGTGGTGGACCACTAAGTTTCTTCTTAGAGATGCTCTGAGCCAA	22455						
QY	901	ACTTATGATGTTATTTGAAGATGTTTTCACCGA	936						
DB	22456	ACTTATGATGTTATTTGAAGATGTTTTCACCGA	22491						

RESULT 4
LOCUS AY096364
DEFINITION Arabidopsis thaliana putative GDP-L-fucose synthetase (At1g17890)
mrna, complete cds.

ACCESSION AY096364
VERSION FLI CDNA
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 1018)
AUTHORS

Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L.,
Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,
Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,
Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Open Reading Frame (ORF) Clones

TITLE JOURNAL
REFERENCE 2 (bases 1 to 1018)
AUTHORS

Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M.,
Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L.,
Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,
Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,
Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission

TITLE JOURNAL
COMMENT

Submitted (16-APR-2002) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
The RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

FEATURES
Source

1..1018
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="U18093"
/note="This clone is in puni 51."
ecotype: Columbia

gene

1..1018
/gene="At1g17890"

CDS

1..987
/gene="At1g17890"
/codon_start=1
/evidence=experimental
/product="putative GDP-L-fucose synthetase"
/protein_id="AAM20005.1"
/db_xref="GI:20465283"

/translation="MADNTGSEMKSGSFMLEKSAKIFVAGHRLVGSATVRKLQDQGF
TNLVRHSELDITQSQSVESFATEKPVQVILAAKVGIIHANNITPADFVGNLQI
QTNVHSAYTHQWDLFLGSSCIYKPAQVPIPEGALLTPELPNEWYATKAGI
KMCQAVRLOHWDALISMPNLVGNDFHPNSHVLPALMRFFHAKKANNADEVVW
GSGSPLEFLHVDLADCVFLMDQYSGFHNVSQVEVTKELAEVLKEVVGFGK
LWMDITPDGTGPKMDSSKSLASLGWTPKISLKGUSQTYEYLENVQKQ"
988..1018
/gene="At1g17890"

3'UTR
BASE COUNT 277 a 214 c 252 g 275 t
ORIGIN

Query Match 71.3%; Score 667.8; DB 8; Length 1018;
Best Local Similarity 82.5%; Pred. No. 3e-186;
Matches 765; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1 ATGCTGTGACAAATCTGCAAAATCTTCGTCGCGGGTCATCGTGTGTTGGTTGGATCTGCC 60
|||||
DB 43 ATGTTAGAGNAATCCCGCAAGATCTTTGTCGTGGACATAGGGGATGTTGGATCTGCC 102
|||||
QY 61 ATTGTCCGCAAGCTTCAGGAACAAGGTTTCACCAATCTCGTTCTTAAACACACGCGCGAG 120
|||||
DB 103 ATTGTCCGCAAGCTTCAGGATCAGGGTTTCACCAATTTGGTCTTTCGAACCCATTCCAGAG 162
|||||
QY 121 CTTGATCTCACTCGTCAAGCGGATGTTGAATCTCTTTCTTCAAGAGAGCCAGTTTAT 180
|||||
DB 163 CTTGATCTCACTAGCCAATCAGATGTTGAATCTTTCTTCTACAGAGAAACCAAGTTTAT 222
|||||
QY 181 GTAATCTTAGCAGCAGCTAAAGTTGTTGGTATTTCAGCTTAACAACCACTTCTCGCTGAT 240
|||||
DB 223 GTCATCTAGCTGCTGCTAAAGTCGGTGGGATTCATGCTTAACAACATACCCAGCTGAT 282
|||||
QY 241 TTCATTGGTGTCAATCTCCAGATTCAGACCAATGTGATCCACTCTCATATGAGCAGCGT 300
|||||
DB 283 TTCATTGGGTCATCTCCAAATCCAGACCAATGTGATCCACTCTCTTATACACAGCT 342
|||||
QY 301 GTGAAGAAGTTCTTCTCTTGATCATCTGATTAATCCCTAAATTTGCTCTCAGCCA 360
|||||
DB 343 GTGAAGAAATCTCTTCTCTTGGTTCATCTGCAATTTACCCCAAAATTTGCTCTCAGCG 402
|||||
QY 361 ATTCCTGAGTCTGCTTTGTTTAACGAGCATCGCTTGAACCACTAATCAGTGGTATGCTATT 420
|||||
DB 403 ATTCGCAATCTGCGCTTCTTACCGTCTGCTCGAACCGACAAACGAATGGTATGGATC 462
|||||
QY 421 GCTAAGATCGTGGGATTAAGACTTTGTCAGGCTTATAGGATTCAGCAGCGATGGATGCA 480
|||||
DB 463 GCCAAGATCGGGTATTAGATGTCTCAAGCTTACCGGCTTCAGCATCAGTGGGATGCT 522
|||||
QY 481 ATCTCTGGCATGCTACTAATCTCTATGCTCTAATGACAATTTCCACCCGGAGAAATTCCT 540
|||||
DB 523 ATCTCCGGTATGCCACGAATCTGTATGTCAGATGATAATTTCCACCCGGAGAAATTCCT 582
|||||
QY 541 CATGTGCTTCTGCTCTTATGAGGAGGTTCCAGCGGCGAAAGTGAATCGAGCGGAGAA 600
|||||
DB 583 CATGTGCTGCTGCTCTCATGAGAGGTTTCATGAAGCCAAAGCCAAATATGCTGACGAA 642
|||||
QY 601 GTTGTGGTGGGGTACAGTATGCTCGGTTGAGGGAGTTCTTGCATGTTGATGTTGGCT 660
|||||
DB 643 GTTGTGGTGGGAGTGGAGCCGTTGAGGAGTTTTCATGTCGATGTTGGCT 702
|||||
QY 661 GATGCTTGTGTTTCTTGTGATTCGATACAGCGGTTGGAGCATGTTTAACTTGAAGT 720
|||||
DB 703 GATGCTTGTGTTTCTTGTGATTCAGTACAGCGGATTTGAGCATGTTGAATGTAGAACT 762
|||||
QY 721 GGTCAAGAACTGACTATTAGAGATTTGGTGGTGGTCAAGAGAGGTTGCTGTTTGA 780
|||||
DB 763 GGTGTTGAAGTACGATCAAGAAATTTGGCTGAGTTGGTGAAGAAAGTTTGGGTTTAA 822
|||||
QY 781 GCGAAGCTTTGGATGGATTTGCACTAAGCCAGATGGCACACCGAGGAAATTTATGACAGC 840
|||||
DB 823 GGAAGACTTTGTTGGGATACCAACAAACCCAGACGGGACACCAAGGAACTGATGATAGC 882
|||||
QY 841 TCAAGCTCGCTCTTTGGTGGACACCTAAGGTTTCTTCTAGAGATGCTGTGACCCA 900
|||||

Db 883 TCGAAGCTTGCTCTTTGGATGGACACGGAAGATTTGCTTTAAAGATGGCCTCTCCCAA 942
QY 901 ACTTATGATGGTATTTGAAGAATGTT 927
|||||
Db 943 ACTTATGATGGTATTTGAAGAATGTT 969
|||||

RESULT 5
AY063880
LOCUS
DEFINITION
Arabidopsis thaliana putative GDP-L-fucose synthetase (At1g17890)
mrna, complete cds.
ACCESSION
VERSION
AY063880.1 GI:17380847
KEYWORDS
FLI CDNA.
SOURCE
Arabidopsis thaliana.
ORGANISM
Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
1 (bases 1 to 1341)
Yanada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Full Length cDNA Clones

Unpublished
2 (bases 1 to 1341)
Yanada, K., Banh, J., Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D.,
Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C.,
Yanamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H.,
Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A.,
Karlin-Neumann, G., Kawai, J., Kim, C., Koeseema, E., Lam, B., Lin, J.,
Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J.,
Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A.,
Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission

Submitted (19-NOV-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
Riken Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-Length cDNA'); Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada, K., Banh, J.,
Banno, F., Chang, E., Dale, J.M., Goldsmith, A.D., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C.,
Yanamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T.,
Karlin-Neumann, G., Kim, C., Koeseema, E., Lam, B., Lin, J., Meyers, M.C.,
Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A.,
Davis, R.W., Ecker, J.R. and Theologis, A.

Yanada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
contributed equally to this work as PIs.

FEATURES
source

1..1341
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="RAFL08-86-B08 (R18093)"
/note="This clone is a modified pBluescript vector
(Lambda ZAP) as a XhoI/SstI insert."
ecotype: Columbia
1..1341
/gene="At1g17890"
1..146
/gene="At1g17890"

QY	781	GGAAAGCTTGGATGGATTGACCTAAAGCCAGATGGCAACCGAGAACTTATGGACAGC	840
Db	969	CGAAAGCTTGTTTGGATACCAAAACACAGCGGACACCAAGAACTGATGGATAGC	1028
QY	841	TCAAAGCTCGCTCTTCCTGGTGGACACCTAAGCTTTCTCTTACAGATGGCTCTGAGCGAA	900
Db	1029	TCGAAGCTTGGCTCTTCCTGGATGGACACCGAAGATTTGCTTAAAGATGGCTCTCCCA	1088
QY	901	ACTTATGATTTGGTATTTCGAAGAATGTT	927
Db	1089	ACTTATGATTTGGTATTTCGAAGAATGTT	1115
RESULT 6			
LOCUS	AC025908/c		
DEFINITION	Arabidopsis thaliana chromosome 1 clone F20L2, *** SEQUENCING IN PROGRESS ***	13 unordered pieces.	
ACCESSION	AC025908		
VERSION	AC025908.1	GI:7259643	
KEYWORDS	HTG; HTGS PHASE1.		
SOURCE	Arabidopsis thaliana		
ORGANISM	Arabidopsis thaliana		
REFERENCE	1	(bases 1 to 65262)	
AUTHORS	Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Khan, S., Kim, C., Altacfi, H., Bei, Q., Chin, C., Chidou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharly, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Pederspiel, N.A., Theologis, A. and Ecker, J.R.		
TITLE	Genomic sequence for Arabidopsis thaliana BAC F20L2 from chromosome 1		
JOURNAL	Unpublished		
REFERENCE	2	(bases 1 to 65262)	
AUTHORS	Ecker, J.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-MAR-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA		
COMMENT	* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.		
	1	2519: contig of 2519 bp in length	
		gap of unknown length	
	2520	6612: contig of 4093 bp in length	
		gap of unknown length	
	6613	10394: contig of 3782 bp in length	
		gap of unknown length	
	10395	14221: contig of 3827 bp in length	
		gap of unknown length	
	14222	17743: contig of 3522 bp in length	
		gap of unknown length	
	17744	22568: contig of 4825 bp in length	
		gap of unknown length	
	22569	26293: contig of 3725 bp in length	
		gap of unknown length	
	26294	33595: contig of 7302 bp in length	
		gap of unknown length	
	33596	38579: contig of 4984 bp in length	
		gap of unknown length	
	38580	44838: contig of 6259 bp in length	
		gap of unknown length	
	44839	48500: contig of 3662 bp in length	


```

/codon_start=1
/evidence=not_experimental
/protein_id="AAF97258.1"
/db_xref="GI:9665056"
/translation="MDKSIPIKAIPVACVRPDLVDDVTQNTSTIPTMSPVLNMPESA
TSPILMVPRLRTIWPNSKWDGADGPGSTGPIKREASDNTNDTAHTNPAPPPEMVIIP
LTIIPSDSSNYSCDAGAPSTGPKRGKPGSKNSTPTPEPKPKYVDPNSLKYV
SRGNFSEITEAETEGNOEIVDSVMRFDVRRRLCQNHHPEDILITTAGNCTKMGV
KNTRRRIGAPGIVHGDITFYWGEMCLVGLHKSNGYGGIDFFTAASAVEGHAAMCVV
TAGQYDETEGDTLLIYSGQGTGVGNARDQEMKGNLALEASVSKGNDVRVRGVI
HPEHNNKIYIVDGMVLSKFWTVTKGKFKFVRLKVPKPNOPPAYAIWKTVENLRN
HOLIDSROGFILEDLSFGAELLRLVPLANEVDEDDKTIIPEDFYIPSOCHSGMMWTEFH
FRQSLGCONCQRHQMHCNCTVQRNGDLLPYHNNILVCRPLIYECGGSCPCPDHC
PRLVOTGLKHLVFKTRNCGLRSLWDPFRAGTPICEFAGLRKTKKEVEDDDYLF
DTSKIYQRFNWEVELLEDSEWQSEFINLPTQVLISAKSKGNVGRPMHNSCPNV
FWQPISEHNGVILLIIGLFAMKHIIPMTLTYDYGVSERSEDEGLFVCPYLSSK
LWPSSEIYRFLNSKGRAWYKIIYRLASQNVSGSLVDKDEPEKLEBEQMGDGFK
EGLSDVLIDKLEVSDCDRTATSRSHTDVPSQSTHQTQTPESFRTPTICDDDTFVS
VSGISRDVSNLIJPEATETPASVQEKMANTRFSNNVSKNGNDEFFIEDFVGPMDTI
DUTMFRDPDSFDONLLYAWDRDKTQKRSFKRIMDAIKRRRREKEYEQOLAIWFG
DADMGCDLVNDKEQSTSIDSKSQTNVSVSEDESEWEL"
/complement(6878..8287)
/gene="F2H15.2"
/complement(join(6878..6934,7022..8287))
/gene="F2H15.2"
/notes="Contains similarity to female sterile
homeotic-related protein Frg-1 from Mus musculus
gb|AF045462 and contains a bromodomain PF|00439."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97259.1"
/db_xref="GI:9665057"
/translation="WSSSEHISGGGASKTKHKWSSQNRKPMGVRSQERSVPLVSPS
NSFASDDHMLKI SLSSISKLEVRNLKRLKSELDVRSLLIKRFPDPAANPGGSMAS
GVGRSKVTKGCGKSGCHGADGTVOIFKNCSLLTKLMKHSAAWVFNVPDVAKG
LGLHDYHNIVKSPMDLTKVTKLGLSKLSPDLDAEDVELTFNNAILYNPICHDVYRF
ABELLNMEDKWSIEMQYDNLHRRKFPTRDIEFPAPASIAPIVEPLPAIVPSPS
SPPPPPPPVAAPVLENTWREESMTIPEPEVTAPEKAEESEAPVNNRDLTLEE
KRLSEELQDLYDKLETVQIIRKSNPELSQKDEIELDLSLINTLWELRYFTVG
YKESLSKNEARHGFSGSDEASVHNSIQEPPTLVSGTTSRVTESGKL"
/misc_feature
8805..9110
/notes="Similar to LINE element"
/gene="F2H15.3"
/notes="Contains similarity to blue copper-binding protein
III from Arabidopsis thaliana gb|U65650 and contains a
Plastocyanin-like PF|02298 domain."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97260.1"
/db_xref="GI:9665058"
/translation="WAQSSGHVSYVAVTVPIAIVMTVLCLFLANAVTVARRPTTVIG
GDGMDPVPMQMTWARGKTFYAGDIEFYDORFNLIIVNRTGYETCEANVGALEYS
SGDDKIQUNGYNYFITYGTPTBEDTGLKMAIKALAPR"
/11220..12453
/gene="F2H15.4"
/join(11220..11361,11662..11927,12058..12453)
/gene="F2H15.4"
/notes="Identical to beta-tonoplast intrinsic protein
(beta-TIP) from Arabidopsis thaliana gb|AF026275 and
contains a MIP (major intrinsic protein) PF|00230 domain.
ESTs gb|R64952, gb|A1999191 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAF97261.1"
/db_xref="GI:9665059"
/translation="WATSARRAYGFRADETHPDPSIRATLAEFLSTFVFPVAGRGSI
LALDKLYMDTAHTGNTNPGVLVALAHALFAAVSAAINVSCHVNPVPAETALI
GGRISVIRAIYVWQALGAILACILLRLATNGLRPVGFHVASGVSELHGLMEIILT
FALVYVWSTALDKRGSGITAPLAIGLIVGANILVGGFPDQASMNPARFAPGALVG
WRWSNHWIYVWGPFGGALALYIYMTIIPSVNEPPHSTHQPAPEDY"

```

```

Query Match 71.3%; Score 667.8; DB 8; Length 75133;
Best Local Similarity 82.5%; Pred. No. 5.1e-186;
Matches 765; Conservative 0; Mismatches 162; Indels 0; Gaps 0;
QY 1 ATGTCGACAAATCTGCCAAATCTTCGCGGGGTCTATCGTGTGTCGATCTGCC 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 36433 ATGTTAGAGAAATCCGCCAAGATCTTTGTCGTGGACATAGGGGATTGTTGTCGATCTGCC 36374

```

QY 61 ATTGTCGCGAGCTTACAGGAACAGGTTTACCAATCTCGTTCTTAAACACACGCGAG 120
Db 36373 ATTGTCGCGAGCTTACAGGAACAGGTTTACCAATCTCGTTCTTAAACACACGAG 36314
QY 121 CTTGATCTCACTGCTCAAGCGGATGTTGAATCTCTTTCTCAAGAGAGCGAGTTTAT 180
Db 36313 CTTGATCTCACTGCTCAAGCGGATGTTGAATCTCTTTCTCAAGAGAGCGAGTTTAT 36254
QY 181 GTAATCTAGCAGCAGCTAAAGTTGGTGTATTCACGCTAACACCTATCTGCTGAT 240
Db 36253 GTAATCTAGCAGCAGCTAAAGTTGGTGTATTCACGCTAACACCTATCTGCTGAT 36194
QY 241 TTCATCTGCTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCTATATGACGCGT 300
Db 36193 TTCATCTGCTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCTATATGACGCGT 36134
QY 301 GTGAAGAGCTTCTCTTCTTGATCATCTGATTAACCTTAATCTGCTCTCAGCCA 360
Db 36133 GTGAAGAGCTTCTCTTCTTGATCATCTGATTAACCTTAATCTGCTCTCAGCCA 36074
QY 361 ATTCCTGAGCTCTGTTGTTAAACAGATCGCTTGAACCACTAATGAGTGTATGCTAT 420
Db 36073 ATTCCTGAGCTCTGTTGTTAAACAGATCGCTTGAACCACTAATGAGTGTATGCTAT 36014
QY 421 GCTAAGATCGCTGGGATTAAGACTTGTGAGCTTTATAGGATTCAGACGATGGGATGCA 480
Db 36013 GCTAAGATCGCTGGGATTAAGACTTGTGAGCTTTATAGGATTCAGACGATGGGATGCT 35954
QY 481 ATCTGCGGATGCTTAACTTCTATGCTTAAATGACAAATTTCCACCGGAGAAATCT 540
Db 35953 ATCTGCGGATGCTTAACTTCTATGCTTAAATGACAAATTTCCACCGGAGAAATCT 35894
QY 541 CATGTGCTTCTCTTATGAGGAGTTCCAGGCGGAAAGTGAATGAGCGGAGGAA 600
Db 35893 CATGTGCTTCTCTTATGAGGAGTTCCAGGCGGAAAGTGAATGAGCGGAGGAA 35834
QY 601 GTTGTGCTGCGGATCACAGTACGCTTGTGAGGAGTTCTGATGATGATTTGGCT 660
Db 35833 GTTGTGCTGCGGATCACAGTACGCTTGTGAGGAGTTCTGATGATGATTTGGCT 35774
QY 661 GATGCTGCTTCTTCTGCTGATTCGATACAGCGGTTGGAGATTTAACTTGAAT 720
Db 35773 GATGCTGCTTCTTCTGCTGATTCGATACAGCGGTTGGAGATTTAACTTGAAT 35714
QY 721 GGTCAAGAGTGAATTAAGAGTTGGCTGAGTTCGTAAGAGGTTCTGCTGTTGA 780
Db 35713 GGTCAAGAGTGAATTAAGAGTTGGCTGAGTTCGTAAGAGGTTCTGCTGTTGA 35654
QY 781 GGAAGCTTGGATGGATTTGCACTAAGCCAGATGGCACACCGAGGAAACTTATGACAGC 840
Db 35653 GGAAGCTTGGATGGATTTGCACTAAGCCAGATGGCACACCGAGGAAACTTATGACAGC 35594
QY 841 TCAAGCTCGCTTCTTGGTGGATGACACCTAAGTTTCTTATGAGATGCTTGAAGCAA 900
Db 35593 TCAAGCTCGCTTCTTGGTGGATGACACCTAAGTTTCTTATGAGATGCTTGAAGCAA 35534
QY 901 ACTTATGATGGTATTTGAAGATGTT 927
Db 35533 ACTTATGATGGTATTTGAAGATGTT 35507

RESULT 8
AY089031
LOCUS
DEFINITION
Arabidopsis thaliana clone 1353 bp mRNA linear PLN 21-JUN-2002
ACCESSION
AY089031
VERSION
AY089031.1
KEYWORDS
FLI CDNA.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1353)
Haas, B.J., Volfovsky, N., Town, C.D., Trukhan, M., Alexandrov, N.,
Feldmann, K.A., Flavell, R.B., White, O., and Salzberg, S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. (2002) In press
2 (bases 1 to 1353)
Brover, V., Trukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and
Feldmann, K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1353)
Brover, V., Trukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and
Feldmann, K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the Ws or Laer ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genset carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.
Location/Qualifiers
1..1353
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="19576"
BASE COUNT 371 a 288 c 293 g 401 t
ORIGIN
Query Match 70.4%; Score 659; DB 8; Length 1353;
Best Local Similarity 82.6%; Pred. No. 1.2e-183;
Matches 766; Conservative 0; Mismatches 160; Indels 1; Gaps 1;
QY 1 ATGCTGCAAAATCTGCCAAATCTTCCTCGGGTCAATCGTGGTTGGTGGATCTGCC 60
Db 199 ATGTTAGAGAAATCCGCCAAATCTTTGTTGCTGCCATAGGGATTTGGTATCCGCC 258
QY 61 ATGTCGCGAAGCTTCAGAACAGGTTTCAACCAATCTCTTCTTAAACACACGCGAG 120
Db 259 ATAGTCCGCAAGCTTCAGGATCAGGTTTCAACCAATCTCTGCTTCAAGAGAACATTTCCGAG 318
QY 121 CTTGATCTCACTCGTCAAGCGGATGTTGAATCTCTTCTTCTCAAGAGAACAGGATTTAT 180
Db 319 CTTGATCTCACTAGCAATCAGATGTTGAATCTCTTCTTCTCAAGAGAACAGGATTTAT 378
QY 181 GTAATCTAGCAGCAGCTAAAGTTGGTGTATTCACGCTAAACACACTATCTGCTGAT 240
Db 379 GTCATCTCGCTCGGGCTTAAAGTCGGTGGAAATTCATGCCAACACACATATCCAGCTGAT 438
QY 241 TTCATCTGCTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCTATATGACGCGT 300
Db 439 TTCATCTGCTCAATCTCCAGATTCAGACCAATGTGATCCACTCTGCTATATGACGCGT 498
QY 301 GTGAAGAGCTTCTCTTCTTGATCATCTGCTTAACTTACCCCTTAAATTTGCTCTCAGCCA 360
Db 499 GTGAAGAGCTTCTCTTCTTGATCATCTGCTTAACTTACCCCTTAAATTTGCTCTCAGCG 558
QY 361 ATTCCTGAGCTCTGTTGTTAAACAGATCGCTTGAACCACTAATGAGTGTATGCTATT 420
Db 559 ATTCCTGAGCTCTGTTGTTAAACAGATCGCTTGAACCACTAATGAGTGTATGCTATT 618

QY 421 GCTAAGATCGCTGGGATTAAGACTTGTACGGCTTATAGGATTCAGACCGATGGGATGCA 480
 |||||
 Db 619 GCCAAGATCGCGGTATTAAAGATGTCTCAAGCTTACCGGCTTCAGCATCAGTGGGATGCT 678
 |||||
 QY 481 ATCTCTGGGATGCTACTATCTCTATGTCTTAATGACAATTTCCACCGGAGATTCT 540
 |||||
 Db 679 ATCTCGGATGCGCCAGCAATCTGTATGTGTGAGATGAATTTCCACCGGAGAAATCC 738
 |||||
 QY 541 CATGTCTTCTCTCTATGAGGAGTTCCAGGCGGAAAGTGAATGAGCGGAGGAA 600
 |||||
 Db 739 CATGTCTGCTCTCTCATGAGAGGTTTCATGAAGCTTAAGCCAAATATGCTGACGA 798
 |||||
 QY 601 GTTGTGGTGGGATACAGTAGTCCGTTGAGGAGTCTTGTCAATTTGATATTGGCT 660
 |||||
 Db 799 GTTGTGGTTGGGAAAGTGAAGCCGTTGAGGAGTTTTTACATGTGATGATTGGCT 858
 |||||
 QY 661 GATGCTTGTGTTTCTTGTGATCGATACAGCGGTTGGAGCATGTTAACATTTGAAGT 720
 |||||
 Db 859 GATGCTTGTGTTTCTTGTGATCAGTACAGCGGATTTGAGCATGTTGAATGTAGGAAGT 918
 |||||
 QY 721 GGTCAAGAAAGTACTATTAGAGAGTTGGCTGAGTTGGTGAAGAGGTTGTTGTTTGA 780
 |||||
 Db 919 GGTGTTGAAGTACGATCAAGAAATTTGGCTGAGTTGTTGTAAGAGTGTGTTGGTTAA 978
 |||||
 QY 781 GGAAGCTTGGATGGATTTGCACTAAGCCAGATGGCACACCGAGGAAACTTATGGACAC 840
 |||||
 Db 979 GGAAGCTTGTGTTGGGATACCAAAACCCAGACCGGACCAAGGAAACTGATGGATAAC 1038
 |||||
 QY 841 TCAAGCTCGCTCTTGGTGGACACCTTAAGTTTCTTAGAGATGCTTGAGCCAA 900
 |||||
 Db 1039 TCGAAGCTTGGCTGTC-TTGGATGGACACCAAGATTTTCGTAAGAGATGCGCTCTCCCA 1097
 |||||
 QY 901 ACTTATGATGGTATTGGAAGATGTT 927
 |||||
 Db 1098 ACTTATGATGGTATTGGAAGATGTT 1124
 |||||

RESULT 9
 AP003565
 LOCUS
 DEFINITION
 Oryza sativa (japonica cultivar-group) chromosome 6 clone
 OSJNBa0085J13, ** SEQUENCING IN PROGRESS **, in ordered pieces.
 ACCESSION
 AP003565
 VERSION
 AP003565.1 GI:13936405
 KEYWORDS
 HTG; HTGS PHASE2.
 SOURCE
 Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
 clone:OSJNBa0085J13
 ORGANISM
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE
 1
 Sasaki.T., Matsumoto.T. and Yamamoto.K.
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, BAC
 clone:OSJNBa0085J13
 JOURNAL
 Published Only in Database (2001)
 2 (bases 1 to 130393)
 Sasaki.T., Matsumoto.T. and Yamamoto.K.
 Direct Submission
 JOURNAL
 Submitted (02-MAY-2001) Takuji Sasaki, National Institute of
 Agrobiological Resources, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced.

* by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source
 1. .130393
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="6"
 /clone="OSJNBa0085J13"
 BASE COUNT 37240 a 26888 c 27585 g 38530 t 150 others
 ORIGIN

Query Match 46.0%; Score 430.6; DB 2; Length 130393;
 Best Local Similarity 67.4%; Pred. No. 8.6e-116;
 Matches 623; Conservative 0; Mismatches 299; Indels 3; Gaps 1;

QY 5 CTGCAAAATCTCCAAAATCTTCGTCGGGGTCATCGTGGTTGGTGGATCGCCATTCG 64
 |||||
 Db 22387 CCGACAAGGCGCGCAAGGTGTTCTGCGCGGCACCGCGGCTCTGTCGGCTCGCCATCC 22446
 |||||
 QY 65 TCCGCAAGCTTCAGGAACAAGGTTTCAACCAATCTCGTTCTTAAACACACGCGGAGCTTG 124
 |||||
 Db 22447 TCGGCCACTCTCTCCCTCGGCTTCAACAAGTGTGTCGCGCACCCACGCGAGCTCG 22506
 |||||
 QY 125 ATCTCACTGCTCAAGCCGATGTTGAATCTCTTTTCTCAAGAGAAGCCAGTTTATGTA 184
 |||||
 Db 22507 ACCTCACCGCCAGTCCGACGTCGAGGCTTCTTCGCGCCGAGCTCCCGCTACGTCG 22566
 |||||
 QY 185 TCTTAGCAGCAGTAAAGTTGGTGTATTACGCTTAACAACACCTATCTCTGTGATTTCA 244
 |||||
 Db 22567 TCTTCGCGCGCCCAAGGTCGCGGCGATCCACGCCAACTTCCACCTTCCGCGCGGAGCTTCA 22626
 |||||
 QY 245 TTGGTGTCAATCTCCAGATTACAGCAATGTGATCCACTCTCATATAGACGCGG--TG 301
 |||||
 Db 22627 TCGCGCCAACTCCAGATCCAGACCAAGCTGTCGACGCCGCGCTCAATGCGGCTCCG 22686
 |||||
 QY 302 TGAAGAAGCTTCTTCTTCTTGGATCATCTGCAATTTACCTAAATTTGCTCTCAAGCCAA 361
 |||||
 Db 22687 TCCGCAAGCTCTCTTCTCTCGGCTCTCTGTCATCTACCCCAAGTTTCGCGCGGAGCCGA 22746
 |||||
 QY 362 TTCTCTGAGTCTGTTGTTTAACAGCATCGCTTGAACCAACTAATGATGGTATGCTATTG 421
 |||||
 Db 22747 TCCCGGAAACTCCCTCTCTCCGCGCCCTCTCGAGCCCAACCAAGTGTGACGCGCTCG 22806
 |||||
 QY 422 TTAAGATCGCTGGGATTAAGACTTGTACGCTTATAGGATTCAGACCGATGGGATGCA 481
 |||||
 Db 22807 CCAAGATCGCGGAATCAAGATGTCCAGGCTTACCGCATCCAGACGCGCTTCGACGCCA 22866
 |||||
 QY 482 TCTTGGCATGCTACTAATCTCTATGTCCTTAATGAACAATTTCCACCGGAGAAATTC 541
 |||||
 Db 22867 TCTCGGCATGCCCCAACCTCTACGGCCCGCAGGACAACTTCCACCGGAGAAATTCAC 22926
 |||||
 QY 542 ATGTCTTCTGCTCTTATGAGGAGTTCACAGGCGGAAAGTGAATGGAGCGGAGGAAG 601
 |||||
 Db 22927 ACGTCTTCCCGCACTGATTCGCGCGGTTTCCATGAGCGGAAGCGTCCAAATGCGCGGAGG 22986
 |||||
 QY 602 TTGTGCTGTGGGTACAGGTAGTCTCGTTGAGGAGTCTTTCATGTTGATGATTTGGCTG 661
 |||||
 Db 22987 TTGTGCTGTGGGCACTGCTCGCGCTGCTGTAATTTCTGACGTCGATGACCTTGCAG 23046
 |||||
 QY 662 ATGCTTGTGTTTCTTGTGATCATACAGCGGTTGGAGCATGTTAACTTTGGAAGTG 721
 |||||
 Db 23047 ATGCGGTGATCTTCTTTGATGGATCACTACTCGGGCTGGAGCATGTCAATGTAGGAAGCG 23106
 |||||
 QY 722 GTCAGAAGTGAATATTAGAGATTGGCTGAGTTGGTGAAGAGGTTGTTGGTTTGAAG 781
 |||||
 Db 23107 GGAGGAGTGAAGTCAAGGAGCTCGCGGAGCTGGTGAAGGAGGTTGGTGGTTCCAGG 23166
 |||||
 QY 782 GGAAGCTTGGATGGATTTGCACTAAGCCAGATGGCACCCGAGGAACTTATGGACAGCT 841
 |||||
 Db 23167 GGAAGCTGTTGGGACTTACGAACCCGATGGCACCCGAGGAGCTCATGGATAGTT 23226
 |||||
 QY 842 CAAAGCTCGCGCTTTTGGGTTGGACACCTAAGGTTTCTTTAGAGATGCTCTGAGCCAAA 901
 |||||

Db 23227 CCAAGTACAGGAGATGGGTGGAAGCCCAAGGTCCCTCAAGAGGGGCTCGTCCAGA 23286
QY 902 CTTATGATTGGTATTGAAGAATGT 926
Db 23287 CATAAAGTGGTATGTCGAGAATGT 23311

RESULT 10
AX172657
LOCUS AX172657 1490 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 147 from Patent WO0144476.
ACCESSION AX172657
VERSION AX172657.1 GI:14597769
KEYWORDS
SOURCE
ORGANISM
Phycomitrella patens.
Phycomitrella patens.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE
AUTHORS Lerchl, J., Renz, A., Ehrhardt, T., Reindl, A., Cirpus, P., Bischoff, P.,
Frank, M., Freund, A., Duwenig, E., Schmidt, R. M. and Reski, R.
TITLE Moss genes from phycomitrella patens encoding proteins involved in
the synthesis of carbohydrates
JOURNAL Patent: WO 0144476-A 147 21-JUN-2001;
BASF Plant Science GmbH (DE)
FEATURES
Location/Qualifiers
source 1. .1490
/organism="Phycomitrella patens"
/db_xref="taxon:3218"
347. .1279
/notes="c_pp001064043r"
/codon_start=1
/protein_id="CAC43708.1"
/db_xref="GI:14597770"

BASE COUNT 369 a 308 c 410 g 402 t 1 others
ORIGIN

Query Match 41.8%; Score 391; DB 6; Length 1490;
Best Local Similarity 65.0%; Pred. No. 2.6e-104;
Matches 596; Conservative 0; Mismatches 315; Indels 6; Gaps 1;

QY 7 GACAAATCTGCCAAATCTTCGCGGGTCAATCGTGGTTGGTGGATCTGCCATTGTC 66
Db 356 GACAAAGGACGCCAAGATCTTTGTTGCTGCACACCGAGGTCTAGTAGTGGCTGTGTT 415
QY 67 CGCAAGCTTCAGGAACAAGTTTCCACCATCTGTTCTTAAACACACCGCCAGCTTGT 126
Db 416 CGTGTCTTGAAGAAGATGGTTATACAAATTTGGTATGAAGACTCATAAAGAACTAGAT 475
QY 127 CTCACCTCGTCAAGCCCATGTTGAATCTTCTTCTTCAAGAGAAGCCAGTTTATGTAATC 186
Db 476 CTTACCGGTTCAGCACAGS-----AATTTTCGACACGGAACACAGCTACGTGATC 529
QY 187 CTAGACAGAGCTAAAGTTGGTGGTATTCACGCTAACACACCTATCTCTGCTGATTTCAAT 246
Db 530 CTAGACAGCTCGAAGGTGGGAGCATTCACGCAACACAGTACTTACCTCGACAGTTCAAT 589
QY 247 GGTGCAATCTCCAGATTTCAGACCAATGATCCACTCTGCATATGACGACCGGTGGAAG 306
Db 590 GCGTGAATCTGCAGATCAAAACGAATGTCATCGATGCTGCTTCAAGAGTCTGGGTGAAG 649
QY 307 AAGCTTCTCTCTCTTGGATCATCTCTGATTTACCTTAAATTTGGCTCTCCTCAGCAATTCCT 366
Db 650 AAGCTCTTCTTCTGGCTCTTGGTGTATCTACCCAAAGTTGGCCAGGTACCCATCGTT 709
QY 367 GAGTCTGCTTTGTTAAACAGCATCGCTTGAAACCAACTAATGATGGTGTATGCTATTGCTAAG 426

Db 710 GAGGAGTCCGCTCCTCAGACGGGCTTTGGAAGCTACAAACGAGTGTGTCTAGCAAG 769
QY 427 ATCGCTGGGATTAAAGACTTTGTCAGGCTTATAGGATTACAGCGGATTCGATCAATCTCT 486
Db 770 ATTGAGGAATCAAAATGTGCCAGGCTTACAGGCTGCAGTATAAATTTGACGCCATTTCT 829
QY 487 GGCATGCTTACTTAATCTCTATGGTCTTAATGACAAATTTCCACCCGGAGAAATTTCTCATGTG 546
Db 830 GGAATGCCGACAAACCTTACGGTCCCCACAGCAATTTCCATCCCGAGAACTCCACGCTC 889
QY 547 CTTCTGCTCTTATGAGGAGTTCCACGAGGCGAAAGTGAATGAGCGGAGGAAGTTGTG 606
Db 890 TTGCCAGGCTTGATCAGACGCTTTTCCAGGAGTAAAGTGAACGCGGCTAAGGAAGTGGTT 949
QY 607 GTGTGGGTACAGGTAGTCCGTTGAGGAGTCTTTCGATGTTGATGATTGGCTATGCT 666
Db 950 GTGTGGGATCAGGTTCCCATTTCCGTCAGTTTCTTTCAGTGGAGCATTTGGCAGAGGCA 1009
QY 667 TGTGTTTTCTTGTCTGATCGATACACGCGGTTGGAGCATGTTAAACATTGGAAAGTGTCAA 726
Db 1010 ACAGTATTTCTGCTGCAAAATTTACTCCGCGCATGAGCATGTCACATGGGCACTGCTCT 1069
QY 727 GAAGTGACTATTAGAGATTGGTGGTGGTGAAGAGGTTGTTGGTTTGAAGGGAG 786
Db 1070 GAGGTCTCAATCAAGGAACTCGCCGAAATGGTGAAGGAAGTGGTTGGATTTTCAGGGG 1129
QY 787 CTTGATGGGATTCACCTAAGCCAGATGGCACACCGGAGAACTTATGGACAGCTCAAG 846
Db 1130 CTGACATGGGATCTTCTTAAGCCTGATGGAATCCACGAAAGCTCATCGATAGCAGCAA 1189
QY 847 CTCGGCTCTTTGGGTTGGACACCTAAGGTTTCTCTTTAGAGATGGTCTCAGCCAACTTAT 906
Db 1190 CTTGCCACATGGGTGGCAAGGAGAAATTCCTCAAGGAAGGATTGGCAGAGACTTAC 1249
QY 907 GATTGGTATTGGAAGA 923
Db 1250 AATGGTACTGTGAGAA 1266

RESULT 11
AE013290/c
LOCUS AE013290 10029 bp DNA linear BCF 17-MAY-2002
DEFINITION Methanosarcina mazei strain Goel, section 72 of 379 of the complete
genome.
ACCESSION AE013290 AE008384
VERSION AE013290.1 GI:20905046
KEYWORDS
SOURCE
ORGANISM
Methanosarcina mazei Goel.
Methanosarcina mazei Goel
Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
Methanosarcinaceae; Methanosarcina.
REFERENCE
AUTHORS Deppe, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A.,
Martinez-Arias, R., Henne, A., Wier, A., Baeumer, S., Jacobi, C.,
Brueggemann, H., Lienard, T., Christmann, A., Boemcke, M., Steckel, S.,
Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P.,
Gunsalus, R. P., Fritz, H.-J. and Gottschalk, G.
TITLE The genome of Methanosarcina mazei: Evidence for lateral gene
transfer between Bacteria and Archaea
J. Mol. Microbiol. Biotechnol. (2002) In press
REFERENCE
AUTHORS Deppe, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A.,
Martinez-Arias, R., Henne, A., Wier, A., Baeumer, S., Jacobi, C.,
Brueggemann, H., Lienard, T., Christmann, A., Boemcke, M., Steckel, S.,
Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P.,
Gunsalus, R. P., Fritz, H.-J. and Gottschalk, G.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute
for Microbiology and Genetics, Grisebachstrasse 8, Goettingen
37077, Germany
FEATURES
Location/Qualifiers
source 1. .10029

```

/organism="Methanosarcina mazei Goel"
/strain="Goel"
/db_xref="taxon:192952"
complement (157. .903)
/gene="MM0653"
complement (157. .903)
/gene="MM0653"
/EC_number="2.4.1.-"
/codon_start=1
/transl_table=11
/product="glycosyltransferase involved in cell wall biogenesis"
/protein_id="AAM30349.1"
/db_xref="GI:20905047"
/translation="MGPKVSVITVCYNABEEFIENAIYSVLNOSYENIYIIIDGASKD
NTVSIINKRYSVMACFISEPDKGYEAMNKKAAATGDILYFLNSDDVFYDYVYVENV
KVFQKNVVELIYGPFIIRDPTNESIKPHDYITKSVFI NSICQOQIFVKASVEN
KCGQNDTVKIVGDYEWELRAFYKNIKKRYEGIIAIFRDGMCNCDKFSDLHAKER
TKVIKEYNIFVIYKYKELLPCRSIAKRLK"
complement (918. .2030)
/gene="MM0654"
complement (918. .2030)
/gene="MM0654"
/EC_number="2.4.1.-"
/codon_start=1
/transl_table=11
/product="mannosyltransferase"
/protein_id="AAM30350.1"
/db_xref="GI:20905048"
/translation="WNILYDYSIFHLQRYGGSRYFYELINRLINLKEASINVPEGFH
INEYLSGNKNWNLKGYKIPIKIYNVVKLDPIMFMTYSHMGNLDIFPHTYIRE
DUDKFRNAAVITVHMDIHLYKQDFKYDPTTKYKRLSVNAADAIICVSENKXKDI
RIYNPBEKIVVYHSGSLNKIKSLMDKFTKENSLLKPKPYLLYVGNRBDGKYNFSLMN
VVAHFADDFNLFCFGGNKFKETELILTKNLLDKVHVNGSDDFLAALYKKAFLCI
YPSLYEGGFCIPLEAMSGLCPVIASNTSSIPREVCGNGGILPDLSEEDLTRSDLSLM
HRSRLILKGGQEQEKPSWDKTASETFNVKSIIS"
complement (2006. .2236)
/gene="MM0655"
complement (2006. .2236)
/gene="MM0655"
/codon_start=1
/transl_table=11
/product="hypochemical protein"
/protein_id="AAM30351.1"
/db_xref="GI:20905049"
/translation="WNHTKYSFYHVVCLFKEIYLLRAYDFVLKLTVHEULTNELNL
AMSRIICCNFKLIFRLYAKPIYRSRIFDSKTS"
complement (2567. .3832)
/gene="MM0656"
complement (2567. .3832)
/gene="MM0656"
/codon_start=1
/transl_table=11
/product="Polysaccharide ABC transporter, ATP-binding protein"
/protein_id="AAM30352.1"
/db_xref="GI:20905050"
/translation="MTASKKERPIIEVHKLSKOYNIGVDVTYKLTSESFTSAVKOPIK
TLKDFRQNETFWALKDYNFEVERGEVLGICRNGAGKSTLLKILSRITSPQGEVRM
RGVGSLLVEVGFPELPSGRNIPFNGSILKKKKEIDEKFEIYKSGVEKFLDTP
VSHNGAVQNLQCORILLNEGLITNGKTADVINTYLTQLQVADKSIKRDVSKKL
FTFDINLNTKEIKTSDENILAIYNLNLVOLTGAQITFEIDQNGICILSSTN
PDSNLIKTLQSPGPIYATCPINVDFLRNLGLYIITISSVNVNIBILDIVEPATRFEI
ESPLDITLLGQSRKGVIFKFWQINQY"
complement (3833. .4681)
/gene="MM0657"
complement (3833. .4681)
/gene="MM0657"
/codon_start=1
/transl_table=11
/product="Polysaccharide ABC transporter, permease protein"

```

```

/protein_id="AAM30353.1"
/db_xref="GI:20905051"
/translation="MCSMTHEATYELVIRPKYGLLDLNMWOLKEYRELLFFLALRE
IKIRYKQITMGASWALQPFMTIIVFTLIFGRLAQMPDGIPIPIFYSGLLLWTYFS
NALSQSSNLVDNAPLLSKVMPRIFFITAPCLSGLDYGIAMSLAIMVYRVFVPG
ISILLPLIILLTFLMAGSGVGLSSICVKYRDVXFILPFFIQLLLFVSPVLYPANIV
GENQLWLLYLNPLTGLISAHRACLLGHIPVDVFGUGISAVLTVVIVFVSGILYLRSTK
YFADLI"
complement (4788. .5726)
/gene="MM0658"
complement (4788. .5726)
/gene="MM0658"
/EC_number="5.1.3.-"
/codon_start=1
/transl_table=11
/product="GDP-fucose synthetase"
/protein_id="AAM30354.1"
/db_xref="GI:20905052"
/translation="MDKRSKIYVAGHRLVGSALRKLKESKDYSNLIIFRTHRELDLTN
QOAVNEFFEREKPEYVFLAAKVGGLIANNTPAEFIYENLMIEANIHASVKCGVKK
LLFLGSSCIYPKLAPQPLKEEYLLTGPLTEETNEAVAYAKIAGIRLCKHYQOYGTNFI
SVMPNLYGPNDFLDETSHVMPALVRFHFAKVNNEPEVITWGTGKPYRFLHVDMM
ADACVYLMENENTDDIGRFVNIQVGKDIITIGELAEILKEIVGFKGEIRKDLSPDGP
QKLDDITKLSSLGHWKANISLXDGIRQTYEWYQSQIK"
complement (5744. .6799)
/gene="MM0659"
complement (5744. .6799)
/gene="MM0659"
/EC_number="4.2.1.47"
/codon_start=1
/transl_table=11
/product="GDP-mannose 4,6 dehydratase"
/protein_id="AAM30355.1"
/db_xref="GI:20905053"
/translation="MTKTALITGITQDGAFLAELNKGVVVHGIKRSSESNFNTARV
DHLXDPHERNVNFMHYGDLTDSNLIIRIQEVQPDDEIYNLAASHVQVQSPETPEYT
ANSDDLGLTLRLLEAIRILGMEKTKLYQASTSEVDKIDIPKETTTPYKSPYAAA
KLAYAITVYNYREAYFIACNGILFNHESPIRGFTVTRKIMTAATKIKHGLEKLYL
GNLDAKRWGFAKDYEAAMLILQOEPYEDVVIATGETHSVREFTLAFKVEIGIEVLW
KKGSGDEVGIDSGTKVLVEVDPRYYRTEVDLLIGDPSKAKEKLGWPKVKLEELVK
IMITADEKAVQENHTEF"
complement (6890. .8197)
/gene="MM0660"
complement (6890. .8197)
/gene="MM0660"
/EC_number="5.3.1.8"
/EC_number="2.7.7.22"
/EC_number="mannose-1-phosphate guanylyl transferase (GDP)"
/codon_start=1
/transl_table=11
/product="mannose-6-phosphate isomerase"
/protein_id="AAM30356.1"
/db_xref="GI:20905054"
/translation="MYRQPKLFGETSLFQETVLRCLVSDISEVVFVVTNEAQKPFVI
GOIKBAGYSIPPENILIEPEGKNTLPAIFGKMEIEKKPGDSIVGVFSSDHLVDRRAAM
ATISSAENLADYLVTFGVVTFPHTYGYIKASEACSPGYRVSEFRKPDPTAKKY
IEEGLNMGMLPETRFLPFPEARAKHAPSVYCGFEKGINOIYECVDNISIDYIME
KSDKVAVYRLDHKNSDLGNFAAIYDELEKDSVGNVHDCDMLNSDGLNLYSRGCKI
VSLIDVKMNVVDISDALLICPKSSSQKVEIYSELKGRKDERAFIGOTVYVPWSYTF
LRAGQKHLRSLNPKLPLEIIEVQUGELVDEEDIVRFDVYVGRR"
complement (8579. .9328)
/gene="MM0661"
complement (8579. .9328)
/gene="MM0661"
/EC_number="2.1.1.-"
/codon_start=1
/transl_table=11
/product="methyltransferase"
/protein_id="AAM30357.1"
/db_xref="GI:20905055"
/translation="MDANEKKEVIKKRFOETIAMSOGTCCSGGGCCGDSSSDISRAIG
YSEADIEAVPDANMGLGCNFPASTAELKPGDIVLDLGSAGDFCLAAQKVGNSGKVI

```



```
Db 3210 TGCCTTCAAAATTTGTATGGTATGAGATAAATTTTGATTTTGAATACTTCTCATGTTATGG 3151
Qy 551 CTGCTCTTATGAGAGGTTCCAGAGCGGAAGTGAATGAGCGGAGGAGTTGTGGTGT 610
Db 3150 CGGCATTAATAAGAGGTTTCATGAGCAAGGTTCCGGGTCTCAGGAATAATCCATAT 3091
Qy 611 GGGGTACAGGTAGTCCGTTGAGGAGTTCCTTGATGTTGATGTTGGCTGATGCTGTG 670
Db 3090 GGGGAAGTGAGAACATATAGAGAAATTTACATATATAGAGGTTTGGCTGATGGAATTA 3031
Qy 671 TTTTCTTGCTGG---ATCGATACAGCGGTTGAGCAATGTTAACTTGGAAAGTGGTCAAG 727
Db 3030 TTTTCTTGATGGAACATGCGCAAAAGTGAAGGTTTAAATATAGGTTGTGGCAAG 2971
Qy 728 AAGTGACTATTAGAGGTTGCTGAGTTGCTGAGGAGGTTGTTGGTTTGAAGGAGC 787
Db 2970 ATATTAAATAAAGATTTAGCTTTATAAAATAAAGATGTTAGTTGATTTAAAGGAAATA 2911
Qy 788 TTGGATGGGATTGCATTAAGCCAGATGGCACACCGAGGAACTTATGGACAGCTCAAAGC 847
Db 2910 TAATTTTGTATAGAGTAAGCCAGATGGAATGTTAGAAAGTTATGGATGTTTCTAAA 2851
Qy 848 TCGGCTCTTGGTTGAGACACCTAAGGTTTCTCTTAGAGATGTTCTGAGCCAACTTATG 907
Db 2850 TTAATAGTTTAGGATGTCATTATAAAGTTGAACTTGATGAAGGAATTAACAAAACATATA 2791
Qy 908 ATTGGTATTGGAAGATGTTT 928
Db 2790 GGTGGTATTAAATAATTGTT 2770
```

RESULT 13

```
A63794
LOCUS A63794 1020 bp DNA linear PAT 12-MAR-1998
DEFINITION Sequence 15 from Patent WO9723624.
ACCESSION A63794
VERSION A63794.1 GI:3717366
KEYWORDS unidentified.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1020)
AUTHORS Hermon-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,
Sumar,N. and Ford,J.
TITLE NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA
AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
JOURNAL Patent: WO 9723624-A 15 03-JUL-1997;
ST GEORGE S HOSPITAL MEDICAL S (GB)
COMMENT Other publication AU 1202797 19970717.
FEATURES
Location/Qualifiers
Source 1..1020
/organism="unidentified"
/db_xref="taxon:32644"
CDS 1..1020
/notes="unnamed protein product; protein sequence is in
conflict with the conceptual translation"
/codon start=1
/protein_id="CAA03745.1"
/db_xref="GI:3717367"
/translation="VRWHTMDRHDVAVLGRSKLTTPGPGLDRATPVYIAGHRLVGS
ALVRFPEAGFTNLIVRSDEIDLTDRAATFDVSETRPQVIIDAAARVVGIMANNY
PADFLSENRIOTNLIDAAVAVRPLFLGSCIVPKYAPQIHESALLTGPLEPTN
DAYATAKIGILQVQAVRQYGLAWISAMPTNLYGPDNFSPSGSHLLPALIRRYEA
KAGAEVNTWGTGTPRELLHVDOLASACIFLEHFDGPNHVNVTGVDSHISIAID
MYATAVCYIGETRNWDTKEDGTPRKLDDVSALRELGRWPRIALKDGDIVTSWYRTNA
DAVR"
```

```
BASE COUNT 207 a 297 c 305 g 211 t
ORIGIN
Query Match 31.7%; Score 297; DB 6; Length 1020;
Best Local Similarity 58.2%; Pred. No. 1.7e-76;
Matches 522; Conservative 0; Mismatches 375; Indels 0; Gaps 0;
```

```
Qy 29 TCGCGGTCATCGTGGTGTGGATCTGCCAATGTCGCAAGCTTTCAGGAACAAGGTT 88
Db 104 TCGCGGTCATCGGCGGCTGGTGGCTCAGCGCTGTAAGTAGATTTGAGCGCGAGGGT 163
Qy 89 TCACCAATCTCGTTCCTTAAACACACGCGAGCTTGATCTCACTCGTCAAGCCGATGTTG 148
Db 164 TCACCAATCTCATGTCGATCAGCGGATGAGATTGATCTGACGGACCGGCGCAACGT 223
Qy 149 AATCCTTCTTTCTCAGAGAAGCCAGTTTATGTAATCTAGCAGCAGCTAAAGTTGGTG 208
Db 224 TTGATTTTGTGCTGAGACAAGCAACAGGTGATCATCGATCGCGCGCAGCGGTCGGC 283
Qy 209 GTATTCACTCACTATCACTTCTGCTGATTTCAATGCTGTCATCTCCAGATTCAGA 268
Db 284 GCATCATGGCGAATTAACACTATCCGCGACTCTCTTGTCCGAAACCTCCGNAATCAGA 343
Qy 269 CCAATGTGATCCACTCTGCATATGAGCAGCGGTGTGAAGAGCTTCTCTTCTTGGATCAT 328
Db 344 CCAATTTGCTCGACGAGCTGTCGCGGTGTCGCGGCTCCTTTTCTCGGTTCTG 403
Qy 329 CTGCAATTTACCTTAATTTGCTCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCAT 388
Db 404 CATGCATCTACCCGAAAGTACGCTCCGCAACCTATCCAGAGAGTGTCTTTATTGACTGGCC 463
Qy 389 CGCTTGAACCAACTAATGAGTGTGATGCTATTGCTAAGATCGCTGGGATTAAGACTTGT 448
Db 464 CTTTGGAGCCCAACCAAGCGGTATCGATCGCCAGATCGCGGTATCTTGCAGATTC 523
Qy 449 AGGCTTATAGGATTCAGCAGCGATGGATGCAATCTTGGCATGCTACTAATCTCTATG 508
Db 524 AGGCGGTTAGCGGCCAATATATGGCTGGCGTGGATCTCTCGATGCGGACTTAACCTCTACG 583
Qy 509 GTCTTAATGACATTTCCACCCGAGAAATTTCTCATGTGCTTCTGCTCTTATGAGAGGT 568
Db 584 GACCCGGGACAACTTCTCCCGTCCGGTCCGATCTCTTCCGCGCTCATCCGTCGAT 643
Qy 569 TCCACGAGCGCAAGTGAATGGAGCGGAGGAATTTGTGTGGGTGAGGTACAGGTAGTCCGT 628
Db 644 ATGAGGAGCCAAAGCTGGTGTGAGAGAGGTACGNAATTTGGGGACCGGTACTCCGC 703
Qy 629 TGAGGAGGTTCTTGATGTTGATGTTGGTGTGATGCTGTGTTGTTTCTTGTGGATTCGAT 688
Db 704 GCGCGAACTTCTGATGTCGACGATCTGGCGAGCGCATGCTGTCTCTTTTGAACAT 763
Qy 689 ACAGCGGTTGAGAGCATCTTAACATTTGAAGTGGTCAAGAAGTACTATAGAGATTGG 748
Db 764 TCGATGGTCCGAAACCAACGTCACCGTGGGCAACCGGCTCGATCACAGCATTTAGCGAGATCG 823
Qy 749 CTGAGTTGGTGAAGAGAGGTTGTGTTTGAAGGGAAGCTTGGATGGGATTCACCTAAGC 808
Db 824 CAGACATGTCGCTACGCGGTGGGCTACATCGCGGAACACAGTTGGGATCCAACTAAC 883
Qy 809 CAGATGGCACACCGAGGAACCTTATGACAGCTCAAGCTCCGCTCTTTGGTTGGACAC 868
Db 884 CCGATGGAACCCCGCGCAAACTATTGGAGCTCTCCGCGCTACCGAGTTGGTTGGCGCC 943
Qy 869 CTAAGTTTCTTTAGAGATGCTCTGAGCCAACTATGATGGTATTGTAAGATG 925
Db 944 CGCGAATCGCACTGAAGACCGCATCGATGCAACCGGTGTCGTGGTACCGCACAAATG 1000
```

RESULT 14

```
MAP223833
LOCUS MAP223833 4435 bp DNA linear BCT 07-JAN-1999
DEFINITION Mycobacterium avium paratuberculosis gs[ba,bb,c,d] genes.
ACCESSION AJ223833
VERSION AJ223833.1 GI:3550477
KEYWORDS gsa gene; gsb gene; gsc gene; gsd gene.
SOURCE Mycobacterium avium subsp. paratuberculosis.
ORGANISM Mycobacterium avium subsp. paratuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
```


QY 869 CTAAGGTTTCTCTTAGAGATGGTCTGAGCCAACTTATGATTGTGTAATTTGAAGAATG 925
 Db 2115 CGCGAATCGCACTCAAGACGGCATCGATCGAACGGTGTCTGTATCCGCGACAAATG 2171

RESULT 15

A63783
 LOCUS A63783 4435 bp DNA linear PAT 12-MAR-1998
 DEFINITION Sequence 4 from Patent WO9723624.
 ACCESSION A63783
 VERSION A63783.1 GI:3717355
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1. (bases 1 to 4435)
 AUTHORS Hermon-Taylor,J., Doran,T., Millar,D., Tizard,M., Loughlin,M.,
 Sumar,N. and Ford,J.
 TITLE NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA
 AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY
 JOURNAL Patent: WO 9723624-A 4 03-JUL-1997;
 ST GEORGE S HOSPITAL MEDICAL S (GB)
 COMMENT Other publication AU 1202797 19970717.
 FEATURES
 Location/Qualifiers
 source 1..4435
 /organism="unidentified"
 /db_xref="taxon:32644"

BASE COUNT 904 a 1289 c 1288 g 954 t
 ORIGIN

Query Match 31.7%; Score 297; DB 6; Length 4435;

Best Local Similarity 58.2%; Pred. No. 2e-76;

Matches 522; Conservative 0; Mismatches 375; Indels 0; Gaps 0;

QY 29 TC CGCGGTATCGTGGTTGGTGGATCTGCCAATGTCGCAAGCTTCGCAAGCAACAAGTT 88
 Db 1275 TC CGCGGTATCGTGGGTGCTGCTGCGCTCAGCGCTCGTAGATTGGAGCGGCGGT 1334
 QY 89 TCACCAATCTGTTCTTAAACACACGCGGCTGATCTCACTCGTCAAGCGGATGTTG 148
 Db 1335 TCACCAATCTGTTGCGGATCAGCGGATGAGATTGATCGCGGACCGGCGCAACGT 1394
 QY 149 AATCCTCTTTCTCAAGAGAAGCCAGTTTATGTAATCCTAGCAGCAGCTAAAGTTGGTG 208
 Db 1395 TTGATTTTGTTGTTGAGACAGACACAGGTGATCATCGATGCGGCGCACGGTGGCG 1454
 QY 209 GTATTCAGCGTAACAACCTATCTCTGCTGATTTCAATTTGGTGTCAATCTCCAGATTCA 268
 Db 1455 GCATCATGGCGAATAACACCTATCCGCGGACTTCTTTCGGAACCTCCGAATCCAGA 1514
 QY 269 CCAATGTGATCCACTCTGCATATGAGCAGGTGTGAAGACTTCTCTTCTTGGATCAT 328
 Db 1515 CCAATTTGCTCGAGCAGCTGTCCCGTGTGCGCGGCTCTTTCTCGGTTCGT 1574
 QY 329 CTGCAATTTACCTAAATTTGCTCTCAGCCAAATTCCTGAGTCTGCTTTGTTAAACAGCAT 388
 Db 1575 CATGCATCTACCCGAAGTAGCTCCGCAACCTATCAAGAGTGTCTTATTCATGGCC 1634
 QY 389 CGCTTGAAACCAACTAATAGTGGTATGCTATTGCTAAGATCGCTGGGATTAAGACTTTGC 448
 Db 1635 CTTTGGAGCCCAACACGACGCGTATGCGATCGCAAGATCGCGGTAATCTCGAAGTTC 1694
 QY 449 AGGCTTATAGGATTCAGACGATGGATGCAATCTCTGGCATGCCCTACTAATCTCTATG 508
 Db 1695 AGGCGTTAGGCGCCCAATATGGGCTGGCTGGATCTCTGCGATGCCGACTAACCTCTACG 1754
 QY 509 GTCTTAATGACATTTCCACCCCGGAGAAATTCATGCTGCTTCTGCTCTTATCAGGAGGT 568
 Db 1755 GACCCGGCGNCAACTTCTCCCGTCCGGGTGCGATCTCTTGGCGGGCTCATCCGTCGAT 1814
 QY 569 TCACACAGGCGAAAGTGAATGAGACGAGGAAATGTTGTGTGGGGTACAGGTAGTCCGT 628

Search completed: June 3, 2003, 03:32:04
 Job time : 2641 secs

Db 1815 ATGAGGAGCCAAAGCTGGTGTGTCAGAAAGAGGTGACGAATTTGGGGACCGGTACTCCGC 1874
 QY 629 TGAGGGAGTTCTTTGCATGTTGATGATTTTGGCTGATGCTTGTGTTTCTTCTGGATCGAT 688
 Db 1875 GGCGGAATCTTGCATGTCGACGATCTGGCAGGCGATGCCTGTCTTTTGGAAACATT 1934
 QY 689 ACAGCGGTTTGGAGCATGTTAAACATTTGGAAGTGGTCAAGAAGTGACTATTAGAGAGTTGG 748
 Db 1935 TCGATGGTCCGAACCCACGTCACGTCGCGGTGCGATCACAGCATTTAGCGAGATCG 1994
 QY 749 CTGAGTTGGTGAAGAGGTTGTTGTTTGAAGGGAAGCTTGGATGGGATTTGCACATAAGC 808
 Db 1995 CAGACATGGTTCGTACGCGGCTGCGGTACATCGCGGAAACACGTTGGGATCCAACTAAAC 2054
 QY 809 CAGATGGCACACCGAGGAAACTTATGGACAGCTCAAAGCTCGCGTCTTTTGGTTGGACAC 868
 Db 2055 CCGATGGAAACCCCGCGCAAACTATTGGACGCTCTCCGCGTACGCGAGTTGGTTGGCGCC 2114
 QY 869 CTAAGGTTTCTCTTAGAGATGGTCTGAGCCAAACTTATGATTGGTATTTTGAAGAATG 925
 Db 2115 CGCGAATCGCACTGAAAGACGCGCATCGATCGCAACGGTGTGCTGTACCGCACAAATG 2171

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 3, 2003, 05:04:31 ; Search time 531 Seconds
(without alignments)
59.476 Million cell updates/sec

Title: US-10-089-014-1
Perfect score: 1646
Sequence: 1 MSDKSAKIFVAGHRLVGS.....SLRDGLSTQDYWLKNVCR 312

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101223694 residues
Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1515	92.0	314	10	US-09-318-271-2
2	1241	75.4	310	10	US-09-734-569-148
3	937.5	57.0	321	10	US-09-318-271-4
4	671	40.8	162	10	US-09-734-569-16
5	463.5	28.2	321	10	US-09-318-271-6
6	249	15.1	311	9	US-09-738-626-3862
7	236	14.3	374	9	US-10-168-066-7
8	215	13.1	420	9	US-10-028-072-260
9	215	13.1	420	9	US-10-121-049-260
10	215	13.1	420	9	US-10-123-904-260
11	215	13.1	420	9	US-10-140-470-260
12	215	13.1	420	9	US-10-175-746-260
13	215	13.1	420	9	US-10-176-918-260
14	215	13.1	420	9	US-10-176-921-260
15	215	13.1	420	9	US-10-227-884-240
16	215	13.1	420	9	US-10-137-865-260
17	215	13.1	420	9	US-10-140-474-260
18	215	13.1	420	9	US-10-142-431-260
19	215	13.1	420	9	US-10-143-114-260

20	215	13.1	420	9	US-10	Sequence 240, App
21	215	13.1	420	9	US-10	Sequence 260, App
22	215	13.1	420	9	US-10-142-419-260	Sequence 260, App
23	215	13.1	420	9	US-10-218-631-240	Sequence 240, App
24	215	13.1	420	9	US-10-230-338-240	Sequence 240, App
25	215	13.1	420	9	US-10-123-262-260	Sequence 260, App
26	215	13.1	420	9	US-10-142-423-260	Sequence 260, App
27	215	13.1	420	9	US-10-230-414-240	Sequence 260, App
28	215	13.1	420	9	US-10-121-050-260	Sequence 260, App
29	215	13.1	420	9	US-10-141-755-260	Sequence 260, App
30	215	13.1	420	9	US-10-143-032-260	Sequence 260, App
31	215	13.1	420	9	US-10-123-108-260	Sequence 260, App
32	215	13.1	420	9	US-10-123-236-260	Sequence 260, App
33	215	13.1	420	9	US-10-123-261-260	Sequence 260, App
34	215	13.1	420	9	US-10-140-921-260	Sequence 260, App
35	215	13.1	420	9	US-10-140-928-260	Sequence 260, App
36	215	13.1	420	9	US-10-216-159A-240	Sequence 240, App
37	215	13.1	420	9	US-10-121-045-260	Sequence 260, App
38	215	13.1	420	9	US-10-123-292-260	Sequence 260, App
39	215	13.1	420	9	US-10-123-903-260	Sequence 260, App
40	215	13.1	420	9	US-10-124-819-260	Sequence 260, App
41	215	13.1	420	9	US-10-124-822-260	Sequence 260, App
42	215	13.1	420	9	US-10-140-925-260	Sequence 260, App
43	215	13.1	420	9	US-10-160-438-260	Sequence 260, App
44	215	13.1	420	9	US-10-218-849-240	Sequence 240, App
45	215	13.1	420	9	US-10-227-873-240	Sequence 240, App

ALIGNMENTS

RESULT 1
US-09-318-271-2
; Sequence 2, Application US/09318271A
; Patent No. US20020012979A1
; GENERAL INFORMATION:
; APPLICANT: Berry, Alan
; APPLICANT: Running, Jeffrey A.
; APPLICANT: Severson, David K.
; APPLICANT: Burlingame, Richard P.
; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"
; FILE REFERENCE: 3161-24
; CURRENT APPLICATION NUMBER: US/09/318,271A
; EARLIER FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 60/125,073
; EARLIER FILING DATE: 1999-03-17
; EARLIER APPLICATION NUMBER: 60/125,054
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 60/088,549
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-318-271-2

Query Match	92.0%	Score 1515;	DB 10;	Length 314;
Best Local Similarity	92.5%	Pred. No. 5.9e-142;		
Matches 298;	Conservative	5;	Mismatches	18;
			Indels	5;
			Gaps	5;
QY	1	MSDKSAKIFVAGHRLVGS	1	MSDKSAKIFVAGHRLVGS
DB	1	MSDKSAKIFVAGHRLVGS	1	MSDKSAKIFVAGHRLVGS
QY	61	VILAAKVGSIHANNTPADFGVNLQIQTNIHSAYEHGVKKLLFLGSSCIYPKFAPOP	120	VILAAKVGSIHANNTPADFGVNLQIQTNIHSAYEHGVKKLLFLGSSCIYPKFAPOP
DB	61	VILAAKVGSIHANNTPADFGVNLQIQTNIHSAYEHGVKKLLFLGSSCIYPKFAPOP	120	VILAAKVGSIHANNTPADFGVNLQIQTNIHSAYEHGVKKLLFLGSSCIYPKFAPOP
QY	121	IPESALLTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLYGPNDFHPENS	180	IPESALLTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLYGPNDFHPENS
DB	121	IPESALLTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLYGPNDFHPENS	180	IPESALLTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMPTNLYGPNDFHPENS

QY 181 HVLPMRRPHEAKVNGAEVVMVGTGS-----PLR-EFLHVDLADACVFLDRYS 231
 Db 180 HVLPMRRPHEAKV-----WSGSCGVGVKVPLEKFLHVDLADACVFLDRIQ 232
 QY 232 -GLEHNVIGSQEVTIRELAELVKEVVEGKLGWCTKPDGTPRKLMDSSKLASLGWTP 290
 Db 233 RGLHNVIGSQEVTIRELAELVKEVVEGKLGWCTKPDGTPRKLMDSSKLASLGWTP 292
 QY 291 KVSRLDGLSQTIDWYLNKNCNR 312
 Db 293 KVSRLDGLSQTIDWYLNKNCNR 314

RESULT 2
 US-09-734-569-148
 ; Sequence 148, Application US/09734569
 ; Patent No. US20020064816A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lerchl, Jens
 ; APPLICANT: Renz, Andreas
 ; APPLICANT: Ehrhardt, Thomas
 ; APPLICANT: Reindl, Andreas
 ; APPLICANT: Cirpus, Petra
 ; APPLICANT: Bischoff, Friedrich
 ; APPLICANT: Frank, Markus
 ; APPLICANT: Freund, Annette
 ; APPLICANT: Duwenig, Elke
 ; APPLICANT: Schmidt, Ralf-Michael
 ; APPLICANT: Reski, Ralf
 ; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
 ; FILE REFERENCE: BASF-NAB-1332-99-US
 ; CURRENT APPLICATION NUMBER: US/09734,569
 ; PRIOR FILING DATE: 2001-05-24
 ; PRIOR APPLICATION NUMBER: US 60/171,101
 ; NUMBER OF SEQ ID NOS: 181
 ; SOFTWARE: Patent in Ver. 2.1/WordPerfect 6.1
 ; SEQ ID NO 148
 ; LENGTH: 310
 ; TYPE: PRT
 ; ORGANISM: Physcomitrella patens
 US-09-734-569-148

Query Match 75.4%; Score 1241; DB 10; Length 310;
 Best Local Similarity 75.5%; Pred. No. 8.4e-115;
 Matches 231; Conservative 37; Mismatches 36; Indels 2; Gaps 1;
 QY 3 DSAKIFVAGHGLVGSATVRKLEQOGFTNLVLTAEHLDTQADVESFSSOEKPVYVI 62
 Db 4 DKAIFVAGHGLVGSATVRKLEQOGFTNLVLTAEHLDTQADVESFSSOEKPVYVI 61
 QY 63 LAAAKVGIIHANNTYPADFIQVNLQIQTNVIHSAYEHGVKLLFLGSSCIYPKFAPQIP 122
 Db 62 LAAAKVGIIHANNTYPADFIQVNLQIQTNVIHSAYEHGVKLLFLGSSCIYPKFAPQIP 121
 QY 123 ESALLTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMTNLYGPNDFHPNSHV 182
 Db 122 ESLLTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMTNLYGPNDFHPNSHV 181
 QY 183 LPALMRPFHAKVNGAEVVMVGTGSPRLRFLHVDLADACVFLDRYSGLHNVIGSQ 242
 Db 182 LPALMRPFHAKVNGAEVVMVGTGSPRLRFLHVDLADACVFLDRYSGLHNVIGSQS 241
 QY 243 EVTIRELAELVKEVVEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKVSRLDGLSQT 302
 Db 242 EVSIRELAELVKEVVEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKVSRLDGLSQT 301
 QY 303 DWYLNK 308
 Db 302 KWCYCN 307

RESULT 3
 US-09-318-271-4
 ; Sequence 4, Application US/09318271A
 ; Patent No. US20020012979A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bert, Alan
 ; APPLICANT: Running, Jeffrey A.
 ; APPLICANT: Severson, David K.
 ; APPLICANT: Burlingame, Richard P.
 ; TITLE OF INVENTION: "VITAMIN C PRODUCTION IN MICROORGANISMS AND PLANTS"
 ; FILE REFERENCE: 3161-24
 ; CURRENT APPLICATION NUMBER: US/09/318,271A
 ; CURRENT FILING DATE: 1999-05-25
 ; EARLIER APPLICATION NUMBER: 60/125,073
 ; EARLIER FILING DATE: 1999-03-17
 ; EARLIER APPLICATION NUMBER: 60/125,054
 ; EARLIER FILING DATE: 1999-03-18
 ; EARLIER APPLICATION NUMBER: 60/088,549
 ; EARLIER FILING DATE: 1998-06-08
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 4
 ; TYPE: PRT
 ; LENGTH: 321
 ; ORGANISM: Escherichia coli
 US-09-318-271-4

Query Match 57.0%; Score 937.5; DB 10; Length 321;
 Best Local Similarity 56.9%; Pred. No. 1.1e-84;
 Matches 177; Conservative 58; Mismatches 67; Indels 9; Gaps 1;
 QY 7 KIFVAGHGLVGSATVRKLEQOGFTNLVLTAEHLDTQADVESFSSOEKPVYVILAAA 66
 Db 5 RVFIAGHGMVGSATVRKLEQOGFTNLVLTAEHLDTQADVESFSSOEKPVYVILAAA 64
 QY 67 KVGGIHANNTYPADFIQVNLQIQTNVIHSAYEHGVKLLFLGSSCIYPKFAPQIPESAL 126
 Db 65 KVGGIHANNTYPADFIQVNLQIQTNVIHSAYEHGVKLLFLGSSCIYPKFAPQIPESAL 124
 QY 127 LTASLEPTNEWYAIKAGIKTCQAYRIQHWDAISGMTNLYGPNDFHPNSHVLPAL 186
 Db 125 LQTLPTNEPYAIKAGIKTCQAYRIQHWDAISGMTNLYGPNDFHPNSHVLPAL 184
 QY 187 MRPFHAKVNGAEVVMVGTGSPRLRFLHVDLADAC-----VFLDRYSGLHVN 237
 Db 185 LRRFHEATQAPADVVVVGSGTMRFLHVDLADAC-----VFLDRYSGLHVN 244
 QY 238 IGSQGVITIRELAELVKEVVEGKLGWCTKPDGTPRKLMDSSKLASLGWTPKVSRLDG 297
 Db 245 VGTGVDCTIRDAVQATKAVGVYGRVVFDAKPDGTPRKLMDSSKLASLGWTPKVSRLDG 304
 QY 298 LSQTYDWYLNK 308
 Db 305 LASTYQWFLN 315

RESULT 4
 US-09-734-569-16
 ; Sequence 16, Application US/09734569
 ; Patent No. US20020064816A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lerchl, Jens
 ; APPLICANT: Renz, Andreas
 ; APPLICANT: Ehrhardt, Thomas
 ; APPLICANT: Reindl, Andreas
 ; APPLICANT: Cirpus, Petra
 ; APPLICANT: Bischoff, Friedrich
 ; APPLICANT: Frank, Markus
 ; APPLICANT: Freund, Annette
 ; APPLICANT: Duwenig, Elke
 ; APPLICANT: Schmidt, Ralf-Michael
 ; APPLICANT: Reski, Ralf

Result No.	Score	Query Match	Length	DB	ID	Description
1	857	91.6	1583	10	US-09-318-271-1	Sequence 1, Appli
2	391	41.8	1490	10	US-09-734-569-147	Sequence 147, App
3	275.2	29.4	966	9	US-09-894-844-49	Sequence 49, Appl
4	269.8	28.8	14187	9	US-10-114-170-121	Sequence 121, App
C 5	265	28.3	536165	9	US-09-939-964-1	Sequence 1, Appli
6	262.2	28.0	966	10	US-09-318-271-3	Sequence 3, Appli
7	213.6	22.8	701	10	US-09-734-569-15	Sequence 15, Appl
8	145.8	15.6	933	10	US-09-962-805-2	Sequence 2, Appli
9	101.6	10.9	1340	10	US-09-318-271-5	Sequence 5, Appli
C 10	59.8	6.4	282	10	US-09-394-0933-2058	Sequence 2058, Ap
C 11	58.8	6.3	447	9	US-10-025-380-296	Sequence 296, App
C 12	58.8	6.3	447	10	US-09-922-217-296	Sequence 296, App
C 13	58.8	6.3	447	10	US-09-833-263-296	Sequence 296, App
14	44.8	4.8	656	10	US-09-815-343-253	Sequence 253, App
C 15	43.8	4.7	846	10	US-09-770-445-638	Sequence 638, App
16	38.6	4.1	492	10	US-09-920-300A-1521	Sequence 1521, Ap
17	38.6	4.1	492	12	US-10-033-528-1521	Sequence 1521, Ap
18	38.4	4.1	1332	9	US-09-938-842A-190	Sequence 190, App
19	38.4	4.1	413	9	US-10-184-644-156	Sequence 156, App


```

121 CTGATCTCACTCGTCAAGCCGATGTTGAATCTCTTTCTCAAGAGAACCCAGTTTAT 180
169 CTGATCTCACTCGTCAAGCCGATGTTGAATCTCTTTCTCAAGAGAACCCAGTTTAT 228
181 GTAACTCTAGCAGCAGCTAAAGTTGGTGTATTCAGCTAAACACACCTATCTCTGCTGAT 240
229 GTAACTCTAGCAGCAGCTAAAGTTGGTGTATTCAGCTAAACACACCTATCTCTGCTGAT 288
241 TTCAATGGTGTCAATCTCCAGATTCAGACCAATGATGATCACTCTGATATGACAGGCT 300
289 TTCAATGGTGTCAATCTCCAGATTCAGACCAATGATGATCACTCTGATATGACAGGCT 348
301 GTGAAGAAGCTTCTCTCTCTGGATCACTCTGATCACTTACCTAAATTTGCTCTCAGCCA 360
349 GTGAAGAAGCTTCTCTCTCTGGATCACTCTGATCACTTACCTAAATTTGCTCTCAGCCA 408
361 ATTCTGAGTCTGCTTTGTTTAAAGCATCGCTTGAACCACTAATGATGATGATGCTATT 420
409 ATTCTGAGTCTGCTTTGTTTAAAGCATCGCTTGAACCACTAATGATGATGATGCTATT 468
421 GCTAAGATCGTGGATTAAGACTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 480
469 GCTAAGATCGTGGATTAAGACTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 528
481 ATCTCTGCGATGCTTACTTCTATGCTCTATGCTCTAATGACCAATTTCCACCGGAGATCT 540
529 ATCTCTGCGATGCTTACTTCTATGCTCTAATGACCAATTTCCACCGGAGATCT 585
541 CATGTGCTTCTGCTTCTTATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 599
586 CATGTGCTTCTGCTTCTTATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 645
600 AGTGTGCTGCTGCTTCTTATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 654
646 AGTGTGCTGCTGCTTCTTATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 705
655 TTGCTGATGCTTCTGCTTCTTATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 711
706 TTGCTGATGCTTCTGCTTCTTATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 765
712 ATTGGAAGTGGTCAAGAGTGAATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 771
766 ATTGGAAGTGGTCAAGAGTGAATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 825
772 GGTGTTTGAAGGAGCTTGAATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 831
826 GGTGTTTGAAGGAGCTTGAATGAGAGGTTCCACGAGCGGAAAGTGAATGAGCGGAGGA 885
832 ATGACAGCTCAAGCTCGGCTCTTCTGCTTGAAGAGTGAATGAGAGGTTCCACGAGCGGAGGA 891
886 ATGACAGCTCAAGCTCGGCTCTTCTGCTTGAAGAGTGAATGAGAGGTTCCACGAGCGGAGGA 945
892 CTGAGCAAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936
946 CTGAGCAAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 990

```

RESULT 2

```

US-09-734-569-147
; Sequence 147, Application US/09734569
; Patent No. US20020064816A1
; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; APPLICANT: Reski, Ralf

```

```

; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; in the synthesis of carbohydrates
; FILE REFERENCE: BASF-NAR-1332-99-US
; CURRENT APPLICATION NUMBER: US/09/734,569
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 60/171,101
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect 6.1
; SEQ ID NO 147
; LENGTH: 1490
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (347)..(1276)
; OTHER INFORMATION: c_dp001064043r
US-09-734-569-147

```

Query Match 41.8%; Score 391; DB 10; Length 1490;

Best Local Similarity 65.0%; Pred. No. 3.6e-119;

Matches 596; Conservative 0; Mismatches 315; Indels 6; Gaps 1;

```

QY 7 GACAAATCTGCCAAATCTTCGTCGCGGTCATCGTGGTTGGTTGGATCTGCCATTGTC 66
DB 356 GACAAAGACGCCAAGATCTTTGTTGCTGGACACCGAGGCTCTAGTAGTGGCGTGT 415
QY 67 CGCAAGCTTCAGGAAACAAAGTTTTCACCAATCTCTGTTCTTAAACACACCGCGAGCTT 126
DB 416 CGTGCTTTGAGNAGGATGGTTATTAACAATTTGGTATGAAGCTCATTAAGAAGTAGAT 475
QY 127 CTCACCTCGTCAAGCCGATGTTGAATCTCTTTCTTCAAGAGAGGCGAGTTTATGTAATC 186
DB 476 CTTACCCGTCAGCACGAGG-----AATTTTCGACACGAGAGAAACGAGGTACGTCATC 529
QY 187 CTAGCAGCAGCTTAAAGTTGGTGGTATTCACGCTTAAACACACCTATCTGCTGATTTCAAT 246
DB 530 CTAGCAGCAGCTTAAAGTTGGTGGTATTCACGCTTAAACACACCTATCTGCTGATTTCAAT 589
QY 247 GGTGTCAATCTCCAGATTCAGACCAATGTATGATCCACTTGCATATGAGCAGCGTGTGAAG 306
DB 590 GCGGTGAATCTGAGATCCAAAGATGTCATGATGCTGCTTCAAGTCTGGGGTGAAG 649
QY 307 AAGCTTCTCTTCTGATGATCTCTGATTTTACCTTAAATTTGCTCTCTCAGCCAAATTCCT 366
DB 650 AAGCTTCTCTTCTGATGATCTCTGATTTTACCTTAAATTTGCTCTCTCAGCCAAATTCCT 709
QY 367 GAGTCTGCTTTGTTAAGCAGCTGCTTGAACCACTAATGAGTGGTATGCTATGCTAAG 426
DB 710 GAGGAGTCTGCTCTGAGCGGCTTTTGAAGGCTTACAAACGAGTGGTATGCTGTAGCAAAG 769
QY 427 ATGCTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGAGTGGGATGCAATCTCT 486
DB 770 ATTGAGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGAGTGGGATGCAATCTCT 829
QY 487 GGCATGCTTCTTATGATGCTTATGATGCTTATGATGCTTATGATGCTTATGATGCTTAT 546
DB 830 GGCATGCTTCTTATGATGCTTATGATGCTTATGATGCTTATGATGCTTATGATGCTTAT 889
QY 547 CTTCTGCTCTTATGAGGAGTTCACGAGGCGGAAAGTGAATGAGGAGGAGGAGTGTG 606
DB 890 TTGCCAGCCTTATGAGGAGTTCACGAGGCGGAAAGTGAATGAGGAGGAGGAGTGTG 949
QY 607 GTGTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGAGTGGGATGCAATCTCT 666
DB 950 GTGTGGGATTAAGACTTGTGAGGCTTATAGGATTCAGACGAGTGGGATGCAATCTCT 1009
QY 667 TGTGTTTCTTCTGATGCTTATGAGGAGTTCACGAGGCGGAAAGTGAATGAGGAGGAG 726
DB 1010 ACAGTATTTCTGCTGAGGATTTACTCCGCGCATGAGCATGTCAACATGCGCAGTGGCTCT 1069
QY 727 GAAGTGAATTTAGAGAGTTCGCTGATGTTGGTGAAGAGGTTGTTGGTTTGAAGGAAG 786

```

5

